

CRITÈRES DE VALIDATION: ASPECTS MÉTHODOLOGIQUES

Julien TOMAS ^{α *} Frédéric PLANCHET ^{α †}

^α ISFA - Laboratoire SAF [‡]

Résumé

Cette note de travail présente les critères quantitatifs et qualitatifs envisagés pour évaluer la pertinence des approches concernant la construction de tables de mortalité. Les divergences entre les approches sont appréhendées sur trois niveaux. Celles-ci concernent l'ajustement de la mortalité passée, la tendance d'évolution de la mortalité ainsi que les intervalles de confiance sur les décès, espérances de vie et autres quantités relatives à la durée de vie ajustée. Les critères de validation sont illustrés à partir des ajustements de la mortalité d'un portefeuille réel obtenus dans la note II1291-12 v1.7. Le choix entre les modèles est partiellement une question de jugement et dépend de l'objectif pour laquelle la table de mortalité prospective est construite. Il appartient aux utilisateurs potentiels de la table de décider des poids à affecter aux différents critères de validation. Néanmoins, le large éventail de critères proposés permet d'apprécier la pertinence des approches.

*Contact: julien.tomas@univ-lyon1.fr

†Contact: frederic.planchet@univ-lyon1.fr.

‡Institut de Science Financière et d'Assurances - Université Claude Bernard Lyon 1 - 50 Avenue Tony Garnier
- 69366 Lyon - France

Sommaire

1 Introduction	1
2 Notations et hypothèses	1
3 Validation des ajustements de la mortalité passée	2
3.1 Ajustement au niveau local	2
3.2 Ajustement au niveau global	2
4 Validation des tendances d'évolution de la mortalité	7
4.1 Marqueurs résumant la distribution de la durée de vie	7
4.2 Plausibilité et cohérence des tendances d'évolution	8
5 Les intervalles de confiances, erreurs relatives et risque systématique d'estimation des paramètres	9
5.1 Intervalles de confiances ponctuels sur les décès et erreur relative	9
5.2 Intervalles de confiances et dispersion relative sur les espérances de vie ajustées	9
6 Application numérique	10
6.1 Validation des ajustements de la mortalité passée	10
6.1.1 Ajustement au niveau local	10
6.1.2 Ajustement au niveau global	14
6.2 Validation des tendances d'évolution de la mortalité	15
6.2.1 Marqueurs résumant la distribution de la durée de vie	15
6.2.2 Plausibilité et cohérence des tendances d'évolution	17
6.3 Les intervalles de confiances, erreurs relatives et risque systématique d'estimation des paramètres	20
6.3.1 Intervalles de confiances ponctuels sur les décès et erreur relative	20
6.3.2 Intervalles de confiances et dispersion relative sur les espérances de vie ajustées	21
7 Conclusion	29
Références	29

1 Introduction

Cette note de travail présente les critères à la fois quantitatifs et qualitatifs envisagés pour évaluer la pertinence des approches concernant la construction de la table de mortalité prospective propre à un organisme.

Nous avons identifié et illustré, dans la note de travail III1291-12 v1.7, quatre approches de positionnement permettant à un organisme donné d'ajuster la référence de mortalité pour se rapprocher d'une vision *best estimate* de ses engagements.

Les divergences entre ces méthodes sont appréhendées sur trois niveaux :

- i. L'ajustement de la mortalité passée : Résidus, tests et quantités résumant les écarts entre l'ajustement et les observations sont utilisés pour comparer les forces et faiblesses de chaque approche.

- ii. Les tendances d'évolution de la mortalité : Il s'agit de confronter les espérances de vie générationnelles partielles et périodiques résiduelles calculées avec les différents modèles et raisonner en terme de plausibilité des tendances d'évolution et de cohérence avec les tendances historiques. En addition, si nous disposons de la mortalité masculine et féminine, nous pouvons comparer les tendances d'amélioration et juger de la plausibilité de l'évolution commune de la mortalité de deux sexes.
- iii. Cette notion est transversale aux deux points précédents. Elle concerne le degré de précision de l'ajustement. Elle est appréhendée par les intervalles de confiances sur les décès, espérances de vie et autres quantités relatives à la durée de vie ajustée.

Le choix entre les modèles est partiellement une question de jugement et dépend de l'objectif pour laquelle la table de mortalité prospective est construite. Il appartiendra aux utilisateurs potentiels de la table de décider des poids à affecter aux différents critères de validation. Néanmoins, le large éventail de critères proposés permettra d'apprécier la pertinence des quatre approches. Nous illustrons les méthodes de validation en Annexes. Les données sont issues d'un portefeuille réel présenté en Annexe de la note de travail III291-12 v1.7.

Cette note de travail est organisée de la manière suivante. La section 2 rappelle de façon précise les notations et hypothèses employées par la suite. La section 3 présente les résidus, tests et quantités permettant d'évaluer l'ajustement de la mortalité passée. La plausibilité et cohérence des tendances d'évolution de la mortalité sont traitées dans la section 4. Les intervalles de confiances et dispersion relative sur les quantités liés à la durée de vie ajustée sont couverts dans la section 5. Enfin, la section 7 présente les conclusions de cette note méthodologique sur les critères de validation.

2 Notations et hypothèses

On note $L_{x,t}$ le nombre d'individus âgés x vivant au 1er janvier de l'année calendaire t . $D_{x,t}$ est le nombre de décès enregistré à l'âge x durant l'année calendaire t d'une exposition au risque $E_{x,t}$.

La probabilité qu'un individu âgé x décède durant l'année calendaire t , notée $\hat{q}_x(t)$, est obtenue à partir de l'estimateur de Hoem,

$$\hat{q}_x(t) = \frac{D_{x,t}}{E_{x,t}},$$

et $\hat{p}_x(t) = 1 - \hat{q}_x(t)$ est la probabilité de survie correspondante.

3 Validation des ajustements de la mortalité passée

3.1 Ajustement au niveau local

Au niveau local, la validation fait intervenir des analyses graphiques. Il s'agit de représenter graphiquement les ajustements contre la mortalité observée pour un âge atteint ou une année calendaire fixe.

En parallèle, il convient d'étudier les graphiques des résidus. On détermine :

Résidus de la réponse. $r_{x,t} = \hat{q}_x(t) - \tilde{q}_x(t)$;

Résidus de Pearson. $r_{x,t} = (D_{x,t} - E_{x,t} \tilde{q}_x(t)) / \sqrt{\text{Var}[E_{x,t} \tilde{q}_x(t)]}$;

Résidus de la déviance. $r_{x,t} = \text{signe}(D_{x,t} - E_{x,t} \tilde{q}_x(t)) \times \sqrt{\text{Déviance}_{x,t}}$.

Les graphiques des résidus fournissent un diagnostic qui complète le graphique des ajustements. Le but est de déterminer si de grands résidus correspondent à des caractéristiques qui ont été mal modélisées localement.

Nous surimposons une courbe *loess*, cf. note III291-11 v1.4, sur les résidus de la réponse et de Pearson afin de visualiser l'écart entre les observations et nos ajustements. En réduisant le bruit, notre attention est plus rapidement portée sur des caractéristiques qui ont été manquées ou ajustées de façon inappropriée. Ici la procédure n'est pas de juger si la courbe surimposée sur les résidus de la réponse est plate. Une droite veut simplement dire qu'aucun manque d'ajustement systématique et reproductible n'a été détecté.

Aucune tendance forte devrait apparaître dans les résidus de la réponse et de Pearson. De plus, si les résidus de Pearson sont dans l'intervalle $[-2; 2]$, cela indique que l'approche modélise correctement la variabilité des données. Enfin, si les résidus de la déviance exhibent plusieurs résidus successifs ayant un même signe, ceci implique que les données sont sur-lissées localement.

3.2 Ajustement au niveau global

Au niveau global, la pertinence de l'ajustement est appréhendée sur deux niveaux :

LA QUALITÉ DU LISSAGE. Il s'agit de déterminer si les données ont été sur-lissées ou sous-lissées. On utilise le **nombre de signes positifs et négatifs** des résidus de la réponse, la valeur du **test des signes** correspondant, le **nombre de runs** et la valeur du **test des runs** correspondant, proposés par Forfar *et al.* (1988, p.56-58) et Debón *et al.* (2006, p.231).

L'ÉCART ENTRE L'AJUSTEMENT ET LES OBSERVATIONS. On utilise le **nombre de résidus standardisés supérieurs à 2 et 3**, et des critères mesurant la distance entre l'ajustement et les observations. On dispose du critère du χ^2 utilisé par Forfar *et al.* (1988, p.56-58) ainsi que le **mean average percentage error** (MAPE) et R^2 utilisés dans Felipe *et al.* (2002). En addition, il nous semble utile de comparer les valeurs du **standardized mortality ratio** (SMR) et de la valeur du **test du SMR** proposé par Liddell (1984), de la **déviance** et de la valeur du **test du rapport de vraisemblance** ainsi que la valeur du **test de Wilcoxon** ou test des rangs appliqué au cas d'échantillons appariés.

Ainsi, l'ajustement est évalué selon la qualité du lissage et l'écart global avec la mortalité passée. Un bon ajustement, marqué par une répartition homogène des signes positifs et négatifs des résidus de la réponse et un grand nombre de *runs* ne doit pas résulter d'un écart trop important avec la mortalité passée, ou réciproquement. En conséquence, on cherchera toujours à équilibrer ces deux aspects complémentaires.

Les tests et quantités résumant la qualité de l'ajustement sont présentés ci dessous :

Test des signes. Il s'agit d'un test non-paramétrique qui examine la fréquence des changements de signes de la différence entre les probabilités de décès observées et ajustées. Sous l'hypothèse nulle \mathcal{H}_0 , la médiane entre les signes positifs et négatifs de cette différence est nulle. Soit le nombre n_+ de signes positifs et n_- de négatifs, avec $n = n_+ + n_-$, la statistique du test des signes, ξ^{SIG} , s'écrit :

$$\xi^{\text{SIG}} = \frac{|n_+ - n_-| - 1}{\sqrt{n}}.$$

Si \mathcal{H}_0 est vraie, cette statistique suit une loi Normale de centrée réduite,

$$\xi^{\text{SIG}} \sim N(0, 1).$$

Ainsi, l'hypothèse nulle \mathcal{H}_0 sera rejetée si

$$|\xi^{\text{SIG}}| > N_{1-\alpha/2}(0, 1),$$

ou $N_{1-\alpha/2}(0, 1)$ est le $(1 - \alpha/2)$ quantile de la distribution Normale centrée réduite. La p-valeur est la plus petite valeur du risque de première espèce (α) pour laquelle on rejette le test. Nous privilégierons donc le modèle ayant la p-valeur la plus proche de 1 :

$$\text{p-valeur} = \mathbb{P}[N_{1-\alpha/2}(0, 1) > |\xi^{\text{SIG}}|] = 2 \times (1 - F_{N(0,1)}(|\xi^{\text{SIG}}|)).$$

Test des runs. Il s'agit d'un test non-paramétrique qui détermine si les éléments d'une séquence sont mutuellement indépendants. Un *run* est le segment d'une séquence, formé par des éléments adjacents égaux. Par exemple, la séquence suivante, composée de 20 éléments,

$$\{ + + - - - + + + - - + + + - - + + \},$$

est composée de 7 *runs* dont 4 sont composés de + et 3 de -. Sous l'hypothèse nulle \mathcal{H}_0 , le nombre de *runs* d'une séquence de n éléments est une variable aléatoire dont la distribution conditionnelle sachant le nombre n_+ de signes positifs et n_- de négatifs, avec $n = n_+ + n_-$ est approximativement Normal, avec :

$$\mu = \frac{2n_+n_-}{n_+ + n_-} + 1 \quad \text{et} \quad \sigma^2 = \frac{2n_+n_-(2n_+n_- - (n_+ + n_-))}{(n_+ + n_-)^2(n_+ + n_- - 1)}.$$

La statistique du test des *runs*, ξ^{RUN} , s'écrit :

$$\xi^{\text{RUN}} = \frac{\text{Nombre de runs} - \mu}{\sigma}.$$

Si \mathcal{H}_0 est vraie, cette statistique suit une loi Normale de centrée réduite,

$$\xi^{\text{RUN}} \sim N(0, 1).$$

Ainsi, l'hypothèse nulle \mathcal{H}_0 sera rejetée si

$$|\xi^{\text{RUN}}| > N_{1-\alpha/2}(0, 1),$$

ou $N_{1-\alpha/2}(0, 1)$ est le $(1 - \alpha/2)$ quantile de la distribution Normale centrée réduite.

La p-valeur est alors donnée par :

$$\text{p-valeur} = \mathbb{P}[N_{1-\alpha/2}(0, 1) > |\xi^{\text{RUN}}|] = 2 \times (1 - F_{N(0,1)}(|\xi^{\text{RUN}}|)).$$

Nous chercherons à avoir la p-valeur la plus proche de 1. Ce test est aussi appelé test de *Wald-Wolfowitz*.

χ^2 . Cet indicateur permet de mesurer la qualité de l'ajustement du modèle. La valeur du χ^2 est

$$\chi^2 = \sum_{(x,t)} \frac{(D_{x,t} - E_{x,t} \tilde{q}_x(t))^2}{E_{x,t} \tilde{q}_x(t) (1 - \tilde{q}_x(t))}.$$

On privilégiera le modèle ayant la valeur du χ^2 la plus faible.

MAPE. Il s'agit d'une mesure de l'exactitude de l'ajustement par rapport aux observations. Cet indicateur correspond à la moyenne des écarts en valeur absolue par rapport aux valeurs observées,

$$\text{MAPE} = \frac{\sum_{(x,t)} |(D_{x,t} / E_{x,t} - \tilde{q}_x(t)) / (D_{x,t} / E_{x,t})|}{\sum_{(x,t)} D_{x,t}} \times 100.$$

C'est donc un pourcentage et par conséquent un indicateur pratique de comparaison. Néanmoins, en présence d'observations nulles il y aura une division par zéro et ces observations doivent être retirées.

R². D'une valeur comprise entre 0 et 1, le coefficient de détermination mesure l'adéquation entre le modèle et les données observées.

Le R² se définit comme la part de variance expliquée par rapport à la variance totale,

$$R^2 = 1 - \left(\frac{\sum_{(x,t)} (D_{x,t} / E_{x,t} - \tilde{q}_x(t))^2}{\sum_{(x,t)} (D_{x,t} / E_{x,t} - (\sum_{(x,t)} (D_{x,t} / E_{x,t}) / n))^2} \right),$$

où n est le nombre d'observations.

La déviance. Il s'agit d'une mesure de qualité de l'ajustement. Sous l'hypothèse des décès suivant une loi de Poisson $D_{x,t} \sim \mathcal{P}(E_{x,t} q_x(t))$, la déviance s'écrit

$$\text{Si } D_{x,t} > 0, \text{ Déviance}_{x,t} = 2 \left(D_{x,t} \ln \left(\frac{D_{x,t}}{E_{x,t} \tilde{q}_x(t)} \right) - (D_{x,t} - E_{x,t} \tilde{q}_x(t)) \right).$$

$$\text{Si } D_{x,t} = 0, \text{ Déviance}_{x,t} = 2 E_{x,t} \tilde{q}_x(t).$$

$$\text{Et Déviance totale} = \sum_{(x,t)} \text{Déviance}_{x,t}.$$

Nous pouvons aussi considérer de calculer la p-valeur associée au test du rapport de vraisemblance (ou *drop-in deviance test*). Nous cherchons ici à déterminer si l'ajustement correspond à la loi de mortalité sous-jacente (hypothèse nulle \mathcal{H}_0).

La statistique du test du rapport de vraisemblance, ξ^{LR} , s'écrit :

$$\xi^{\text{LR}} = \sum_{(x,t)} \left(D_{x,t} \ln \left(\frac{D_{x,t}}{E_{x,t} \tilde{q}_x(t)} \right) - (D_{x,t} - E_{x,t} \tilde{q}_x(t)) \right).$$

Si \mathcal{H}_0 est vraie, cette statistique suit une loi de χ^2 avec un nombre de degrés de liberté égale au nombre d'observations n :

$$\xi^{\text{LR}} \sim \chi^2(n).$$

Ainsi, l'hypothèse nulle \mathcal{H}_0 sera rejetée si

$$\xi^{\text{LR}} > \chi_{1-\alpha}^2(n),$$

ou $\chi_{1-\alpha}^2(n)$ est le $(1 - \alpha)$ quantile de la distribution χ^2 avec n degrés de liberté. La p-valeur est alors donnée par

$$\text{p-valeur} = \mathbb{P}[\chi_{1-\alpha}^2(n) > \xi^{\text{LR}}] = 1 - F_{\chi^2(n)}(\xi^{\text{LR}}).$$

Nous chercherons à avoir la p-valeur la plus proche de 1.

SMR. Il s'agit du rapport entre le nombre de décès observés et ajustés. Si on considère que les décès suivent une loi de Poisson $D_{x,t} \sim \mathcal{P}(E_{x,t} \tilde{q}_x(t))$,

$$\text{SMR} = \frac{\sum_{(x,t)} D_{x,t}}{\sum_{(x,t)} E_{x,t} \tilde{q}_x(t)}.$$

Ainsi, si le $\text{SMR} > 1$ les décès ajustés sont sous-estimés et réciproquement si $\text{SMR} < 1$. Nous pouvons aussi appliquer un test afin de déterminer si le SMR est significativement différent de 1, c.f. **Liddell (1984)**. On calcule la statistique suivante,

$$\begin{aligned} \text{Si } \text{SMR} > 1, \xi^{\text{SMR}} &= 3 \times D^{\frac{1}{2}} (1 - (9D)^{-1} - (D/E)^{\frac{1}{3}}). \\ \text{Si } \text{SMR} < 1, \xi^{\text{SMR}} &= 3 \times D^{*\frac{1}{2}} ((D^*/E)^{\frac{1}{3}} + (9D^*)^{-1} - 1), \end{aligned}$$

où $D = \sum_{(x,t)} D_{x,t}$, $D^* = \sum_{(x,t)} D_{x,t} + 1$ et $E = \sum_{(x,t)} E_{x,t} \tilde{q}_x(t)$.

Si le SMR n'est pas significativement différent de 1 (hypothèse nulle \mathcal{H}_0 vraie), cette statistique suit une loi Normale centrée réduite,

$$\xi^{\text{SMR}} \sim N(0, 1).$$

Ainsi, l'hypothèse nulle \mathcal{H}_0 sera rejetée si

$$\xi^{\text{SMR}} > N_{1-\alpha}(0, 1),$$

ou $N_{1-\alpha}(0, 1)$ est le $(1 - \alpha)$ quantile de la distribution Normale centrée réduite. La p-valeur est alors donnée par

$$\text{p-valeur} = 1 - F_{N(0,1)}(\xi^{\text{SMR}}).$$

Nous chercherons à avoir la p-valeur la plus proche de 1.

Test de Wilcoxon ou test des rangs appliqué au cas d'échantillons appariés. Le cadre de ce test ressemble fortement au test des signes. Alors que le test des signes n'utilise que l'information sur la direction des différences entre les paires composées des probabilités de décès observées et ajustées, le test de Wilcoxon ou *Wilcoxon Matched-Pairs Signed-Ranks test* prend en compte, en plus, la grandeur des différences. On teste l'hypothèse nulle \mathcal{H}_0 que la médiane entre la différence de chaque paire est nulle.

On calcule les différence entre les probabilités de décès observées et ajustées, puis on les classe par ordre croissant des valeurs absolues, en omettant les différence nulles. On affecte à chaque différence non nulle son rang dans le classement (ou la moyenne de ses rangs en cas d'ex-aequo). On note w_+ et w_- la somme des rangs des différences strictement positives et négatives respectivement. Enfin, on note w le plus grand des deux nombres : $w = \max\{w_+, w_-\}$.

Si les probabilités de décès observées et ajustées sont équivalentes, donc si \mathcal{H}_0 est vraie, la somme des rangs ayant un signe positif et celle des rangs ayant un signe négatif devraient être à peu près égale. Mais si la somme des rangs de signes positifs est très différente de celle des rangs de signes négatifs, nous en déduisons que les probabilités de décès ajustées diffèrent de celles observées, et rejèterons l'hypothèse nulle.

Si le nombre d'observations n est supérieur à 15, on calcul la statistique ξ^{WIL} ,

$$\xi^{\text{WIL}} = \frac{(w - 1/2 - n(n+1))/4}{\sqrt{n(n+1)(2n+1)/24}}.$$

Si \mathcal{H}_0 est vraie, cette statistique suit une loi Normale de centrée réduite,

$$\xi^{\text{WIL}} \sim N(0, 1).$$

Ainsi, l'hypothèse nulle \mathcal{H}_0 sera rejetée si

$$|\xi^{\text{WIL}}| > N_{1-\alpha/2}(0, 1),$$

ou $N_{1-\alpha/2}(0, 1)$ est le $(1 - \alpha/2)$ quantile de la distribution Normale centrée réduite. La p-valeur est donnée par :

$$\text{p-valeur} = \mathbb{P}[N_{1-\alpha/2}(0, 1) > |\xi^{\text{WIL}}|] = 2 \times (1 - F_{N(0,1)}(|\xi^{\text{WIL}}|)).$$

On cherche à avoir la p-valeur la plus proche de 1. Le test de Wilcoxon utilise la grandeur des différences. Le résultat peut être différent de celui du test des signes qui utilise le nombre de signes positifs et négatifs de la différence.

4 Validation des tendances d'évolution de la mortalité

Il s'agit dans un premier temps d'obtenir la fonction de survie calculée à partir des tables complétées (avec un modèle de fermeture) résultantes des différents modèles. À partir de la fonction de survie, on peut dériver une série de marqueurs résumant les distributions de la durée de vie. On s'intéresse à la distribution de survie des cohortes d'un âge donné \tilde{x} à la date t . On travaille donc le long de la diagonale du diagramme de Lexis. On peut alors apprécier, en conséquence, les tendances d'évolution de la mortalité et comparer le niveau et la vitesse d'amélioration entre les modèles.

Nous considérons, dans cette section, **l'espérance de vie générationnelle partielle**, **l'espérance de vie périodique résiduelle**, **l'âge médian au décès** ainsi que **l'entropie**. D'autres marqueurs sont possibles, cf. *Pitacco et al. (2009, p.58-63)*, mais ne sont pas développés dans cette note.

Dans un deuxième temps, on se posera la question de la plausibilité et de la cohérence avec les tendances historiques.

Enfin, si nous disposons de la mortalité masculine et féminine, nous pouvons comparer les tendances évolution de la mortalité des deux sexes et juger de la plausibilité de l'amélioration commune.

4.1 Marqueurs résumant la distribution de la durée de vie

Les indices résumant la distribution de la durée de vie sont présentés ci dessous.

La fonction de survie. La fonction de survie d'une cohorte âgée \tilde{x} à la date t mesure la proportion d'individus appartenant à la cohorte âgée \tilde{x} à la date t restant en vie à l'âge $\tilde{x} + x$ (ou de façon équivalente, à la date $t + x$). Sous la condition $S_{\tilde{x}}^{\nearrow}(0) = 1$, elle s'écrit

$$S_{\tilde{x}}^{\nearrow}(x) = \prod_{j=0}^{x-1} (1 - q_{\tilde{x}+j}(t+j)).$$

où l'indice supérieur \nearrow rappelle que l'on travaille le long d'une diagonale du diagramme de Lexis.

Espérance de vie générationnelle partielle. Il s'agit de l'espérance de vie partielle (sur ω années) d'un individu de la cohorte âgée \tilde{x} à la date t . Elle s'écrit,

$$\omega e_{\tilde{x}}^{\nearrow} = \int_1^{\omega} S_{\tilde{x}}^{\nearrow}(u) du.$$

Sous l'hypothèse des forces de mortalité constantes entre les valeurs entières de x et t , on obtient

$${}_{\omega}e_{\tilde{x}}^{\rightarrow} = \sum_{x=1}^{\omega} \prod_{j=0}^{x-1} (1 - q_{\tilde{x}+j}(t+j)) .$$

Espérance de vie périodique résiduelle. Il s'agit de l'espérance de vie résiduelle (sur ω années) d'un individu âgé x à la date t . Elle s'écrit,

$${}_{\omega}e_x^{\uparrow}(t) = \sum_{\delta=1}^{\omega} \prod_{j=0}^{\delta-1} (1 - q_{x+j}(t)) ,$$

où l'indice supérieur \uparrow rappelle que l'on travaille le long d'une verticale du diagramme de Lexis.

L'âge médian au décès. L'âge médian au décès d'un individu de la cohorte âgée \tilde{x} à la date t (sur ω années), notée $\text{Med}[\omega T_{\tilde{x}}]$, est la valeur médiane de la distribution de probabilité de la durée de vie aléatoire $T_{\tilde{x}}$,

$$S_{\tilde{x}}^{\rightarrow}(\text{Med}[\omega T_{\tilde{x}}]) = 0.5 .$$

L'entropie. L'entropie, $H[\omega T_{\tilde{x}}]$, est moins la moyenne de $\ln S_{\tilde{x}}^{\rightarrow}(x)$ pondérée par $S_{\tilde{x}}^{\rightarrow}(x)$ (sur ω années),

$$H[\omega T_{\tilde{x}}] = - \frac{\int_1^{\omega} S_{\tilde{x}}^{\rightarrow}(u) \ln S_{\tilde{x}}^{\rightarrow}(u) du}{\int_1^{\omega} S_{\tilde{x}}^{\rightarrow}(u) du} .$$

Lorsque les décès deviennent plus concentrés, la valeur de $H[\omega T_{\tilde{x}}]$ diminue. En particulier, $H[\omega T_{\tilde{x}}] = 0$ si la fonction de survie a une forme parfaitement rectangulaire.

4.2 Plausibilité et cohérence des tendances d'évolution

On peut raffiner le choix du modèle en appliquant le concept de plausibilité des tendances d'évolution (par exemple en confrontant les espérances de vie générationnelles partielles, ${}_{\omega}e_{\tilde{x}}^{\rightarrow}$, calculées avec les différentes méthodes).

Cette notion, en anglais *biological reasonableness*, a été proposé par Cairns *et al.* (2006) comme un moyen d'évaluer la cohérence des tendances de mortalité extrapolées. Elle n'est basée sur aucun fait scientifique, biologique ou médical. Il s'agit plutôt d'une conception subjective et interroge l'actuaire sur les facteurs biologiques, les progrès médicaux et les changements environnementaux qui expliqueraient cet ensemble particulier de prévisions sachant la provenance des données.

À cela, il convient d'évaluer la cohérence des prévisions avec les tendances historiques. On peut alors comparer les tendances d'évolution des espérances de vie périodiques résiduelles, ${}_{\omega}e_x^{\uparrow}(t)$ à un âge x donné avec les tendances d'évolutions de la mortalité calculées à partir des données brutes.

En addition, si nous disposons de la mortalité masculine et féminine, nous pouvons comparer les tendances d'évolution communes et juger de la plausibilité des évolutions.

On pourra comparer le niveau de la mortalité par âge ainsi que la vitesse d'amélioration. Dans le premier cas, on souhaite s'assurer que la mortalité masculine reste supérieure à celle féminine. Dans le second cas, on vérifie que les taux d'amélioration de la mortalité masculine sont supérieurs à leurs homologues féminins (du moins pour les aux âges élevés).

5 Les intervalles de confiances, erreurs relatives et risque systématique d'estimation des paramètres

Cette notion est transversale aux deux sections précédentes. Elle concerne le degré de précision de l'ajustement. Elle est appréhendée par les intervalles de confiances sur les décès, espérances de vie et autres quantités relatives à la durée de vie ajustée.

5.1 Intervalles de confiances ponctuels sur les décès et erreur relative

Une validation classique est la comparaison entre les décès prédits et observés âge par âge et année par année sur l'ensemble de la période d'observation commune.

En effectuant l'approximation usuelle d'une loi de Poisson par une loi normale, on peut confronter graphiquement les décès observés, prédits ainsi que les bornes inférieure et supérieure de confiance ponctuel à 95 %. En supposant la relation suivante :

$$D_{x,t} \sim \mathcal{N}(E_{x,t} q_x(t), E_{x,t} q_x(t) (1 - q_x(t))),$$

une approximation des $100\% \times (1 - \alpha)$ intervalles de confiance ponctuels pour $D_{x,t}$ est

$$\left(E_{x,t} \tilde{q}_x(t) - z_{1-\alpha/2} \sqrt{E_{x,t} \tilde{q}_x(t) (1 - \tilde{q}_x(t))}, E_{x,t} \tilde{q}_x(t) + z_{1-\alpha/2} \sqrt{E_{x,t} \tilde{q}_x(t) (1 - \tilde{q}_x(t))} \right),$$

ou $z_{1-\alpha/2}$ est le $(1 - \alpha/2)$ quantile de la distribution normale. Les décès observés doivent se situer à l'intérieur des intervalles de confiance théoriques sur l'ensemble de la plage d'âge considérée et sur l'ensemble de la période d'observation commune, signe d'une représentation correcte de la réalité par la table d'expérience.

L'erreur relative est mesurée par la demi-longueur relative de l'intervalle de confiance à 95 %. On obtient ainsi un indicateur par âge et/ou global de l'imprécision de l'ajustement.

5.2 Intervalles de confiances et dispersion relative sur les espérances de vie ajustées

En pratique, l'erreur relative entre les décès théoriques et les décès observés est due à plusieurs sources de risques de nature différente, cf. [Planchet and Thérond \(2011\)](#). Dans cette section, on se consacre uniquement au risque d'estimation des paramètres du modèle, c'est-à-dire le risque systématique associé à l'utilisation du calibrage d'une loi de durée dont on sait qu'il n'est pas exact, conséquence des erreurs d'échantillonnage affectant sa construction, cf. [Kamega and Planchet \(2013\)](#).

On souhaite obtenir les intervalles de confiances ponctuels afin d'encadrer les espérances de vie périodiques résiduelles et mesurer la dispersion relative associée à chaque méthode. La démarche retenue à cet effet consiste, dans un premier temps, à ré-échantillonner les taux de décès bruts à partir de la méthode de simulation directe des taux présentée dans par [Kamega and Planchet \(2013\)](#).

Pour obtenir les taux de décès brutes, on s'appuie sur la simulation du nombre de décès $D_{x,t}^k$ selon

$$D_{x,t} \sim \mathcal{P}(E_{x,t} \tilde{q}_x(t)),$$

par la méthode de simulation de Monte-Carlo.

Pour chaque simulation k et pour chaque approche, on en déduit $q_x^k(t) = D_{x,t}^k / E_{x,t}$ et on estime $\hat{q}_x^k(t)$. Ces différents taux ajustés issus des k simulations représentent la dispersion des taux de décès ajustés au titre du risque systématique sur l'estimation des paramètres compte tenu des fluctuations d'échantillonnage (risque d'estimation).

Pour quantifier cette dispersion, on calcule une approximation des coefficients de variations des taux de décès,

$$cv_{x,t} = \frac{\sqrt{(1/(K-1)) \sum_k (\tilde{q}_x^k(t) - (1/K) \sum_k \tilde{q}_x^k(t))^2}}{(1/K) \sum_k \tilde{q}_x^k(t)}.$$

Enfin, on va chercher à encadrer les espérances de vie périodiques résiduelles. En considérant les taux ajustés issus des simulations, on a donc pour la simulation k :

$$\omega e_x^\uparrow(t)_k = \sum_{\delta=1}^{\omega} \prod_{j=0}^{\delta-1} (1 - q_{x+j}^k(t)),$$

On calcule alors les intervalles de confiance autour de ces espérances de vie périodiques résiduelles,

$$\mathbb{P}[\omega e_x^\uparrow(t)_i < \omega e_x^\uparrow(t)_k \leq \omega e_x^\uparrow(t)_s] = 1 - \alpha,$$

où $\omega e_x^\uparrow(t)_i$ et $\omega e_x^\uparrow(t)_s$ représentent les bornes inférieures et supérieures respectivement et $1 - \alpha$ représente la probabilité que l'espérance de vie périodique résiduelle issue de la simulation k se trouve entre ces bornes. En considérant un seuil α , on obtient les quantiles à $\alpha/2$ et $1 - \alpha/2$ des K simulations d'espérance de vie périodique résiduelle. On peut alors estimer les bornes $\omega e_x^\uparrow(t)_i$ et $\omega e_x^\uparrow(t)_s$,

$$\begin{aligned} \omega e_x^\uparrow(t)_i &= \inf \left\{ \omega e_x^\uparrow(t)_y \in [\omega e_x^\uparrow(t)_1, \dots, \omega e_x^\uparrow(t)_K] \mid \mathbb{P}[\omega e_x^\uparrow(t)_k \leq \omega e_x^\uparrow(t)_y] \geq \alpha/2 \right\} \\ \omega e_x^\uparrow(t)_s &= \inf \left\{ \omega e_x^\uparrow(t)_y \in [\omega e_x^\uparrow(t)_1, \dots, \omega e_x^\uparrow(t)_K] \mid \mathbb{P}[\omega e_x^\uparrow(t)_k \leq \omega e_x^\uparrow(t)_y] \geq 1 - \alpha/2 \right\}, \end{aligned}$$

avec $y \in [1, K]$. Enfin, on peut mesurer la dispersion relative par,

$$\text{dispersion relative} = \frac{95^{\text{ème}} \text{ percentile} - 5^{\text{ème}} \text{ percentile}}{50^{\text{ème}} \text{ percentile}}.$$

6 Application numérique

Les approches théoriques présentée ci-dessus sont illustrés dans cette partie à partir des ajustements de la mortalité d'un portefeuille réel obtenus dans la note II1291-12 v1.7. Pour rappel, il s'agit d'un portefeuille composé uniquement de contrats d'épargne. Les données sont présentées dans la note II1291-12 v1.7, Annexe A, p13.

6.1 Validation des ajustements de la mortalité passée

6.1.1 Ajustement au niveau local

Les Figures 1 et 2 présentent les probabilités de décès et les logarithmes des probabilités de décès obtenues par les quatre méthodes, cf. note II1291-12 v1.7, pour la plage d'âges [30 – 95] sur l'année 2007, pour la population masculine et féminine. Ces graphiques nous donnent une première indication de la qualité de l'ajustement au niveau local.

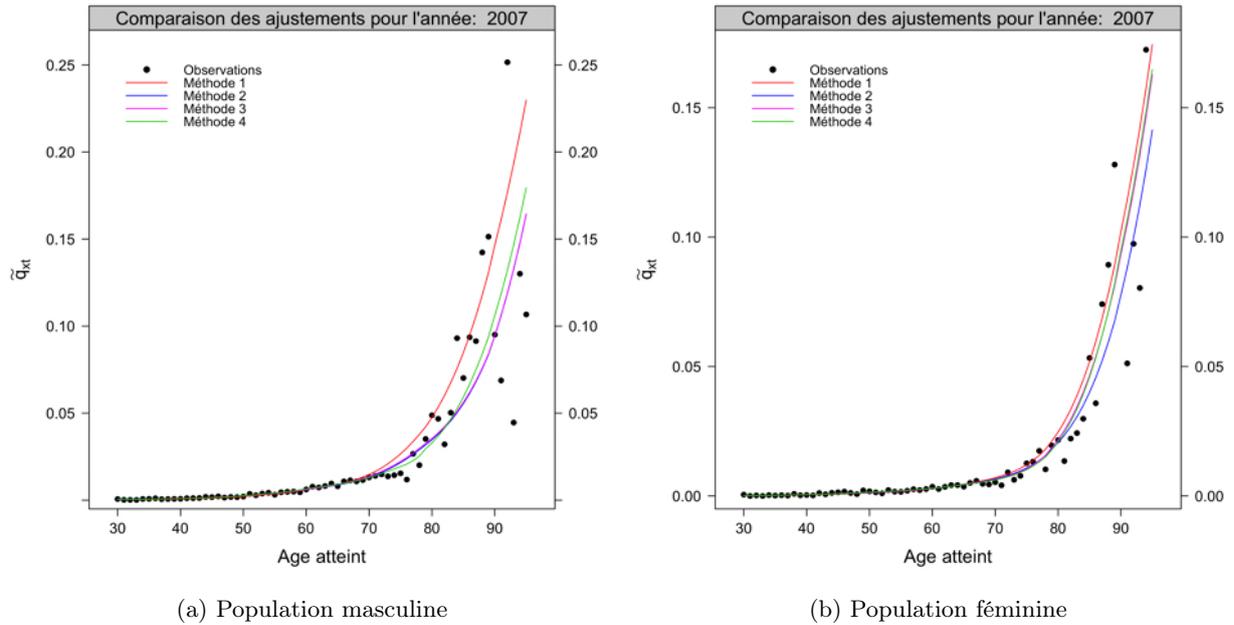


FIGURE 1: Comparaison des probabilités de décès pour l'année 2007.

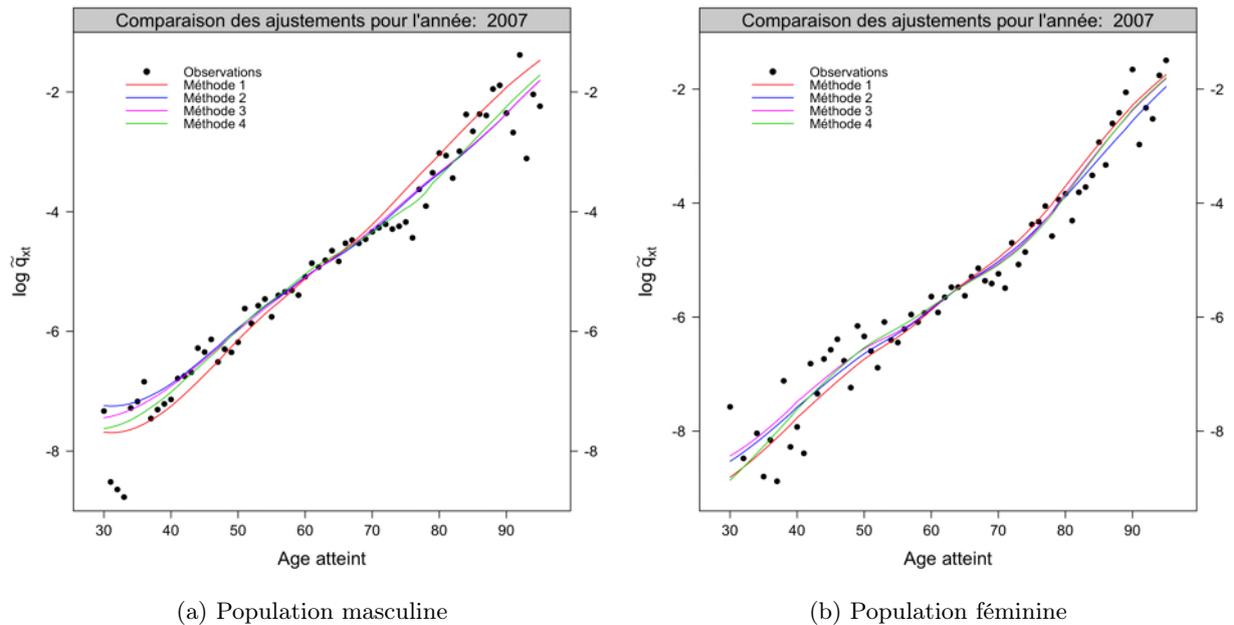
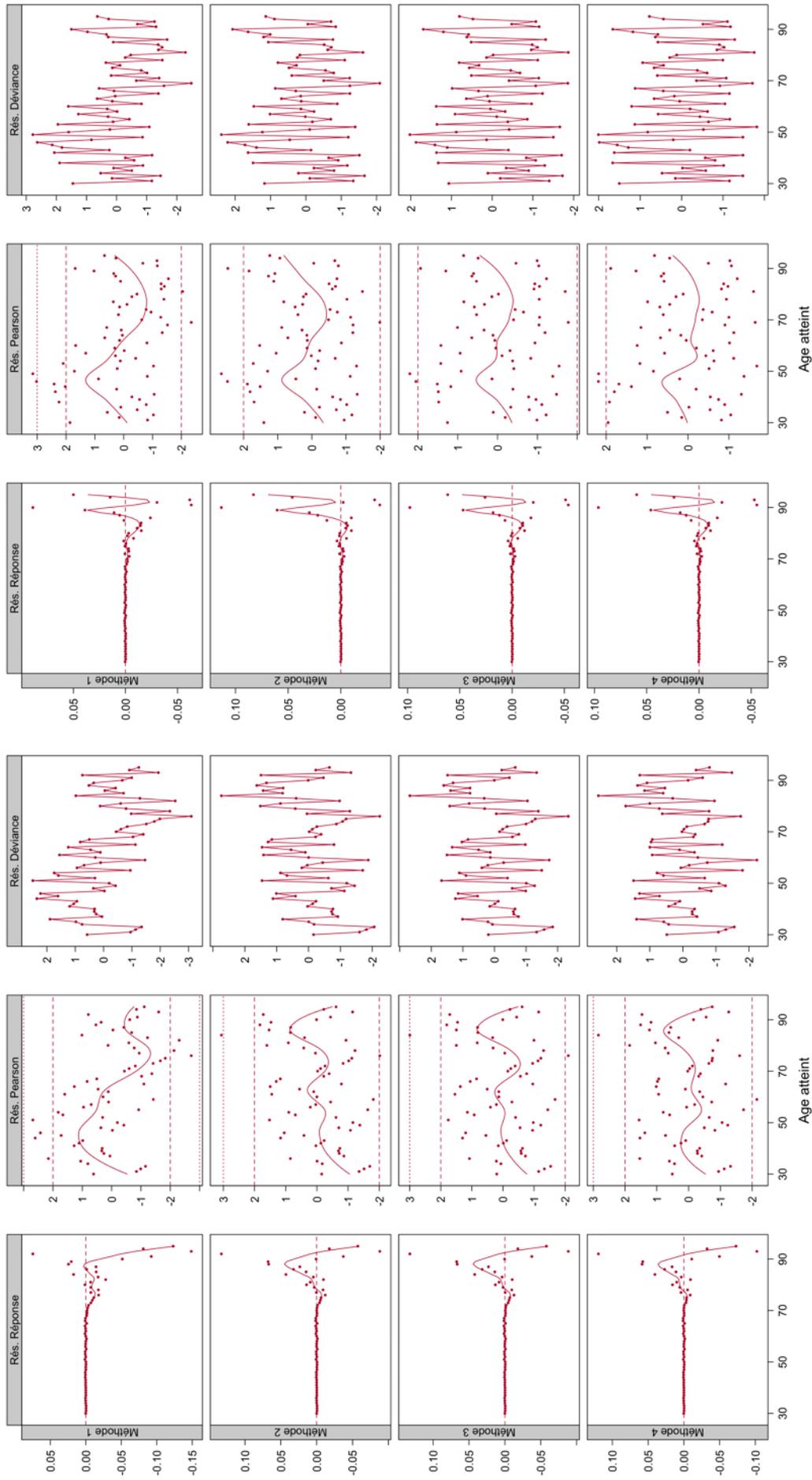


FIGURE 2: Comparaison des probabilités de décès pour l'année 2007, échelle logarithmique.

Les graphiques des résidus des quatre modèles pour l'année 2007 sont présentés en Figures 3a et 3b, pour la population masculine et féminine respectivement.

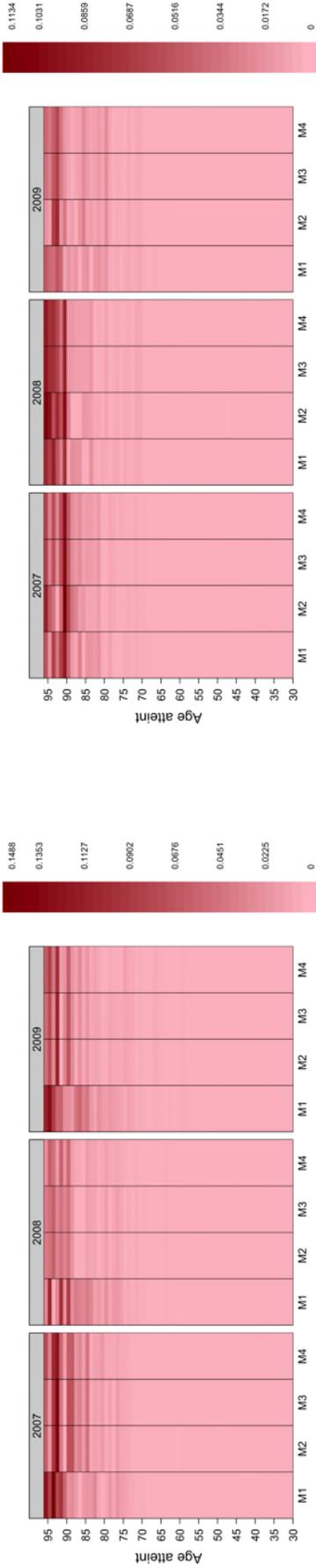
Les résidus de la réponse et de Pearson des quatre méthodes présentent une tendance légèrement plus forte sur la plage d'âges [75 – 90] et [35 – 50], que pour les autres âges, respectivement pour la population masculine et féminine, indiquant un ajustement inapproprié dans ces régions. Cette remarque se retrouve dans les résidus de la déviance. Nous observons que, pour la population masculine, les résidus de la déviance pour les modèles 2, 3, et 4 présentent plusieurs résidus successifs ayant un signe négatif puis positif sur les plages d'âges [71 – 77] et [81 – 89] respectivement. Ceci indique que les modèles sur-estiment les probabilités de décès pour les âges [71 – 77] et les sous-estiment pour les âges [81 – 89].



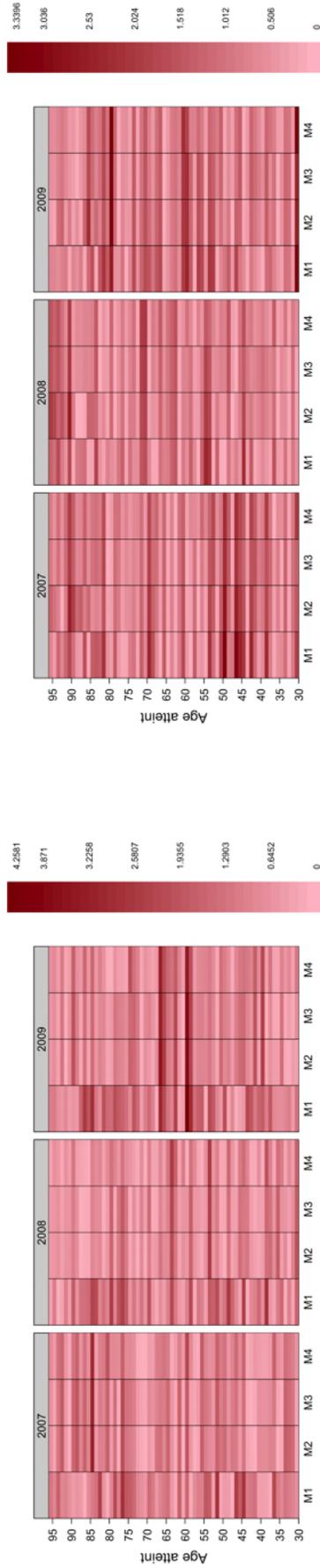
(a) Population masculine

(b) Population féminine

FIGURE 3: Comparaison des résidus pour l'année 2007.

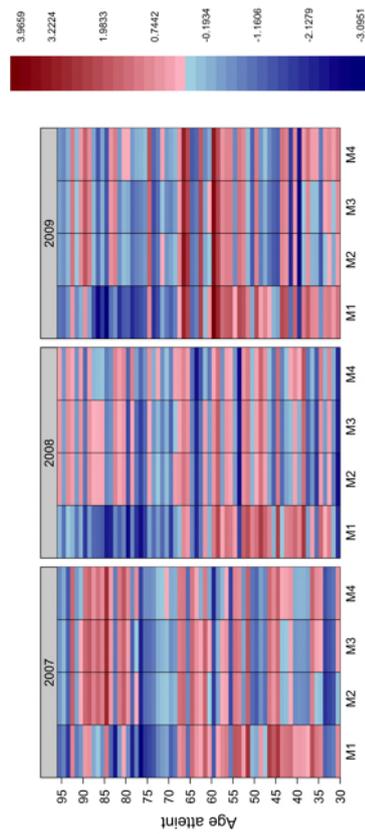


(a) Résidus de la réponse, pop. masculine



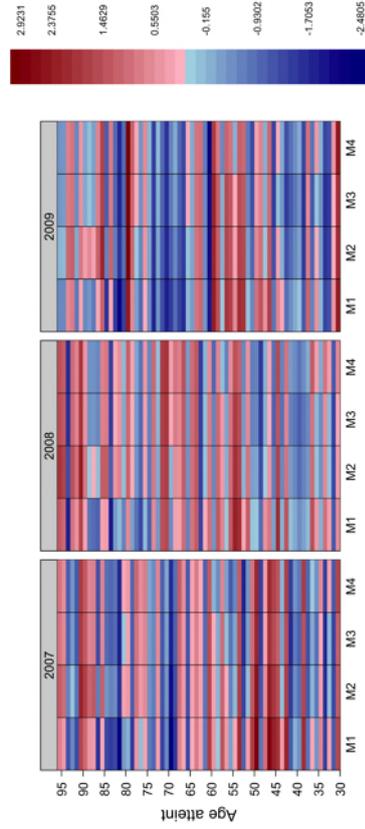
(b) Résidus de la réponse, pop. féminine

(c) Résidus de Pearson, pop. masculine



(e) Résidus de la déviance, pop. masculine

(d) Résidus de Pearson, pop. féminine



(f) Résidus de la déviance, pop. féminine

FIGURE 4: Comparaison des résidus pour les années 2007 à 2009.

Concernant le modèle 1, nous observons plusieurs résidus successifs ayant un signe positif et négatif sur des plages d'âges importantes, $[34 - 47]$ et $[68 - 80]$ montrant clairement un ajustement inapproprié. Pour la population féminine les résidus de la déviance pour les modèles 2, 3, et 4 présentent, au contraire du modèle 1 des résidus qui s'articulent relativement correctement autour de 0. Par ailleurs, les résidus de Pearson sont principalement dans l'intervalle $[-2, 2]$, indiquant que les modèles capturent de façon adéquate la variabilité des données.

Nous pouvons visualiser la qualité de l'ajustement des quatre méthodes pour les années communes avec la référence par les figures 4a à 4f pour la population masculine et féminine respectivement.

6.1.2 Ajustement au niveau global

Nous comparons les probabilités de décès brutes aux séries graduées. Les Tables 1 et 2 présentent les résultats pour la population masculine et féminine respectivement.

		Méthode 1	Méthode 2	Méthode 3	Méthode 4
Test des signes	$+(-)$	92(106)	86(112)	87(111)	96(102)
	ξ^{SIG}	0.9239	1.7767	1.6345	0.3553
	p-valeur	0.3556	0.0756	0.1021	0.7223
Test des runs	Nbre de runs	60	89	86	90
	$ \xi^{\text{RUN}} $	5.6577	1.3476	1.8145	1.4133
	p-valeur	0	0.1778	0.0696	0.1576
Résidus standardisés	> 2	28	10	10	7
	> 3	1	2	2	2
χ^2		354.73	215.07	212.30	193.50
R^2		0.5573	0.7861	0.7873	0.7952
MAPE (%)		38.06	29.61	28.49	26.41
Déviance		369.30	231.680	225.76	202.08
Test du rapport de vraisemblance	ξ^{LR}	356.77	235.95	218.15	194.17
	p-valeur	0	0.0336	0.1554	0.5636
Test du SMR	SMR	1	1.0021	1	0.9998
	ξ^{SMR}	0.0044	0.1591	0.0088	0.0086
	p-valeur	0.4983	0.4368	0.4965	0.4966
Test de Wilcoxon	w	13284	10849	11173	10584
	ξ^{WIL}	4.2523	1.2362	1.6375	0.9079
	p-valeur	0	0.2291	0.1015	0.3639

TABLE 1: Comparaison entre les ajustements, population masculine.

D'une manière générale, les 4 méthodes donnent des résultats acceptables. Néanmoins, on remarque que la qualité de l'ajustement est ordonnée suivant la complexité des modèles.

		Méthode 1	Méthode 2	Méthode 3	Méthode 4
Test des signes	+(-)	97(101)	97(101)	93(105)	95(103)
	ξ^{SIG}	0.2132	0.2132	0.7817	0.4975
	p-valeur	0.8312	0.8312	0.4344	0.6189
Test des <i>runs</i>	Nbre de <i>runs</i>	94	102	96	104
	$ \xi^{\text{RUN}} $	0.8496	0.2132	0.502	0.5940
	p-valeur	0.3956	0.8312	0.6030	0.5525
Résidus standardisés	> 2	15	9	5	5
	> 3	3	1	1	2
χ^2		250.40	226.44	199.00	196.60
R^2		0.8571	0.7750	0.8414	0.8568
MAPE (%)		31.63	30.86	33.00	30.70
Déviance		249.01	226.82	213.59	198.98
Test du rapport de vraisemblance	ξ^{LR}	240.01	300.01	203.19	188.74
	p-valeur	0.0222	0	0.3851	0.6697
Test du SMR	SMR	1	1.0287	1	0.9994
	ξ^{SMR}	0.0061	1.5410	0.0121	0.0210
	p-valeur	0.4976	0.0617	0.4952	0.4916
Test de Wilcoxon	w	10542	10767	9994	9950
	ξ^{WIL}	0.8559	1.1346	0.1771	0.1226
	p-valeur	0.4973	0.2565	0.8594	0.9024

TABLE 2: Comparaison entre les ajustements, population féminine.

Le modèle non-paramétrique endogène, méthode 4, étant le plus flexible, produit les résultats les plus satisfaisants. Que ce soit pour la population masculine et féminine, le modèle présente la plus faible déviance, χ^2 et *MAPE*, ainsi que le nombre le plus faible de résidus standardisés supérieurs à 2. Il exhibe, entre autre, la p-valeur la plus élevée au test du rapport de vraisemblance et de Wilcoxon.

6.2 Validation des tendances d'évolution de la mortalité

Le choix entre les ajustements est affiné en comparant les tendances d'évolution de la mortalité (niveau et la vitesse d'amélioration) entre les modèles. On s'interroge aussi sur la plausibilité et la cohérence avec les tendances historiques observées.

6.2.1 Marqueurs résumant la distribution de la durée de vie

À partir de la fonction de survie, on dérive une série de marqueurs résumant les distributions de la durée de vie. On s'intéresse à la distribution de survie des cohortes de plusieurs âges donnés, i.e. 30 à 80 ans en 2007 durant 50 ans. On travaille donc le long de la diagonale du diagramme de Lexis.

La table 3 présente l'espérance de vie générationnelle partielle, l'âge médian au décès ainsi que l'entropie pour la population masculine et féminine.

	Population masculine				Population féminine			
	Méthode 1	Méthode 2	Méthode 3	Méthode 4	Méthode 1	Méthode 2	Méthode 3	Méthode 4
$\overset{\wedge}{50}e_{30}$	47.71	47.35	47.09	47.63	48.82	48.72	48.89	48.77
$\overset{\wedge}{50}e_{40}$	44.07	44.23	43.58	44.64	46.79	46.89	47.32	46.92
$\overset{\wedge}{50}e_{50}$	36.05	37.58	36.62	37.81	40.69	41.53	42.16	41.22
$\overset{\wedge}{50}e_{60}$	25.89	28.13	27.32	28.13	30.69	32.07	32.34	31.48
$\overset{\wedge}{50}e_{70}$	16.72	19.22	18.68	19.17	20.83	22.34	22.22	21.65
$\overset{\wedge}{50}e_{80}$	9.17	11.71	11.43	11.22	11.90	13.45	12.88	12.52
$Med[{}_{50}T_{50}]$	39.88	42.52	41.09	42.66	44.77	46.36	46.93	45.52
$Med[{}_{50}T_{60}]$	28.75	31.63	30.60	31.65	33.71	35.44	35.57	34.54
$Med[{}_{50}T_{70}]$	18.61	21.49	21.04	21.57	23.15	24.98	24.70	24.01
$Med[{}_{50}T_{80}]$	10.18	13.19	12.88	12.56	13.32	15.17	14.44	14.03
$H[{}_{50}T_{30}]$	0.0009	0.0011	0.0012	0.0010	0.0005	0.0005	0.0004	0.0005
$H[{}_{50}T_{40}]$	0.0028	0.0027	0.0031	0.0025	0.0013	0.0013	0.0011	0.0013
$H[{}_{50}T_{50}]$	0.0111	0.0084	0.0098	0.0083	0.0056	0.0047	0.0042	0.0051
$H[{}_{50}T_{60}]$	0.0490	0.0338	0.0382	0.0350	0.0278	0.0222	0.0215	0.0251
$H[{}_{50}T_{70}]$	0.2179	0.1478	0.1618	0.1552	0.1345	0.1085	0.1096	0.1228
$H[{}_{50}T_{80}]$	1.0274	0.6833	0.7243	0.7417	0.6832	0.5495	0.5814	0.6302

TABLE 3: Indices résumant les durées de vie pour des cohortes de plusieurs âges en 2007 durant 50 années.

On constate que les méthodes 2, 3 et 4 produisent des résultats similaires. A contrario, les indices résumant les durées de vie obtenus par la méthode 1 s'en écartent. D'une manière générale, pour les deux populations, la méthode 1 entraîne les espérances de vie les plus faibles pour des cohortes âgées de 50 ans et plus en 2007. En conséquence, l'âge médian du décès est le plus faible des quatre méthodes et les décès après 50 sont les plus étalés.

La méthode 2 conduit aux espérances de vie les plus grandes pour les cohortes d'âges élevés que ce soit pour la population masculine et féminine. Elle entraîne donc, pour les cohortes âgées de 80 ans, l'âge médian au décès le plus grand ainsi que les décès les plus concentrés. Pour des cohortes plus jeunes, les méthodes 4 et 3 donnent les espérances de vie les plus élevées respectivement pour la population masculine et féminine.

Les Figures 5a à 5h présentent le rapport entre les espérances de vie périodiques résiduelles observées et ajustées respectivement pour la population masculine et féminine et pour les 4 méthodes.

Les espérances de vie périodiques résiduelles issues des données brutes sont de manière générale supérieures à celles calculées à partir des 4 méthodes quelle que soit la population. De plus, les approches sous estiment davantage les espérances de vie périodiques résiduelles féminines que celles masculines.

6.2.2 Plausibilité et cohérence des tendances d'évolution

Dans un premier temps, on compare les tendances d'évolution de la mortalité des deux sexes afin de juger de la plausibilité de l'amélioration commune. Les Figures 6a à 6d comparent le rapport des espérances de vie générationnelles partielles entre les populations féminine et masculine, pour les quatre méthodes respectivement. Les Figures 6e à 6h présentent le rapport entre les probabilités de décès ajustées féminines et masculines.

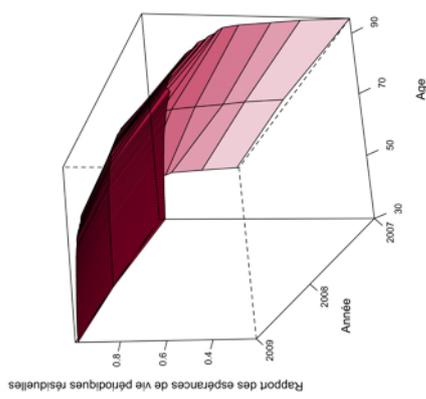
Nous remarquons que les méthodes 1, 2 et 4 donnent des tendances d'évolutions de la mortalité relativement similaires. Les espérances de vie générationnelles partielles féminines sont entre 12 et 15 % supérieures à celles masculines à 110 ans, entre 8 et 10 % à 100 ans et 5 % à 90 ans, cf. Figures 6a, 6b et 6d. Le rapport tend à se rapprocher de 1 avec l'année calendaire, signe que la mortalité masculine s'améliore plus rapidement que la mortalité féminine, ce qui nous semble cohérent.

A l'opposé, le rapport des espérances de vie générationnelles partielles obtenues par la méthode 3 tend à s'éloigner de 1, Figure 6c, indiquant que la mortalité féminine s'améliore plus rapidement. Ces remarques se retrouvent dans les Figures 6a à 6b. De plus, aucune des méthodes induit une mortalité féminine supérieure à celle masculine.

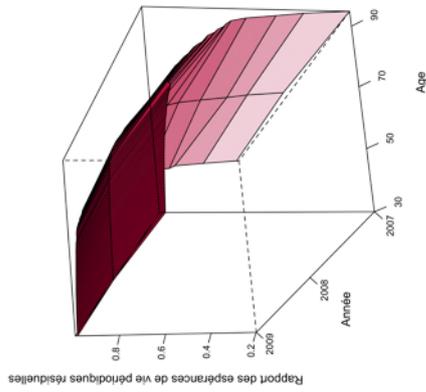
Les figures 7 et 8 comparent les tendances d'évolution des espérances de vie périodiques résiduelles pour les âges 60, 75 et 80 pour la population masculine et féminine respectivement.

On s'aperçoit que quels que soient les âges considérés, les espérances de vie périodiques résiduelles issues des méthodes 2, 3 et 4 suivent convenablement les tendances d'évolutions de la mortalité calculées à partir des données brutes. Seules, celles dérivées de la méthode 1 semblent s'en écarter.

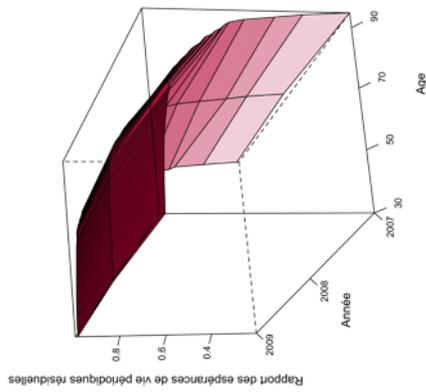
Par ailleurs, les méthodes 2 et 4 donnent des espérances de vie périodiques résiduelles relativement similaires.



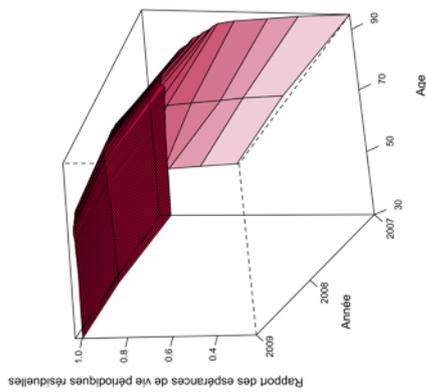
(a) Méthode 1, population masculine



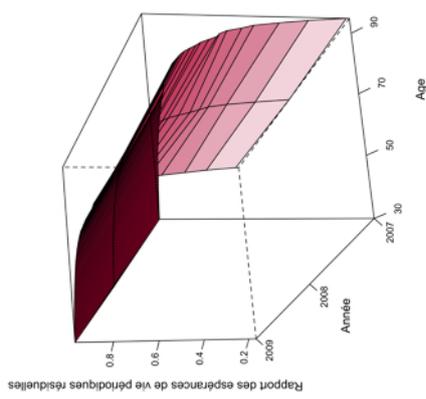
(b) Méthode 2, population masculine



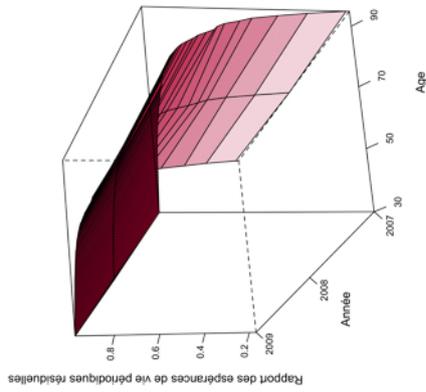
(c) Méthode 3, population masculine



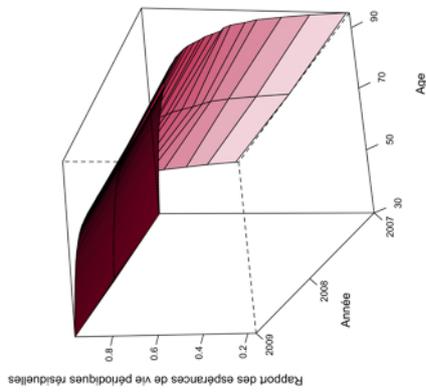
(d) Méthode 4, population masculine



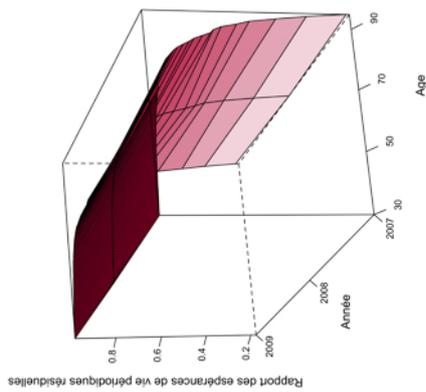
(e) Méthode 1, population féminine



(f) Méthode 2, population féminine

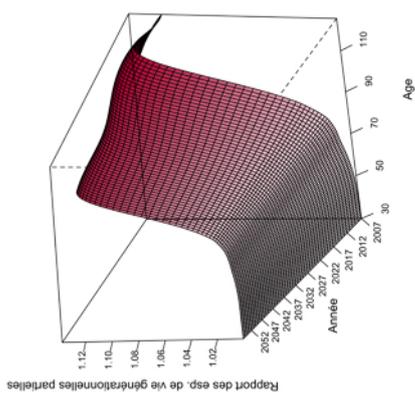


(g) Méthode 3, population féminine

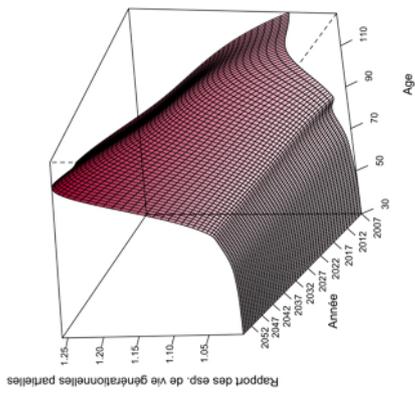


(h) Méthode 4, population féminine

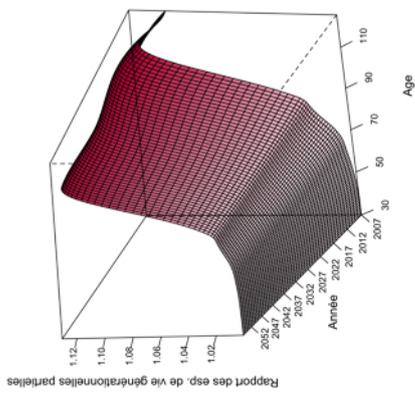
FIGURE 5: Comparaison des espérances de vie périodiques résiduelles observées et ajustées.



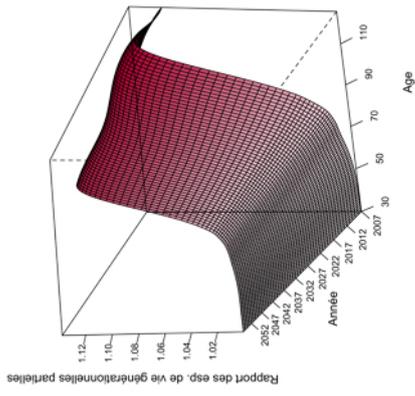
(a) Méthode 1



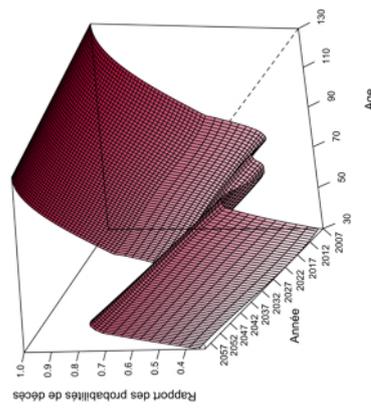
(b) Méthode 2



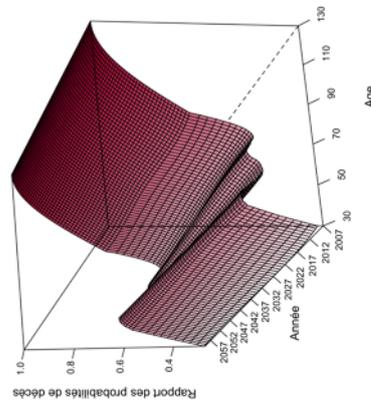
(c) Méthode 3



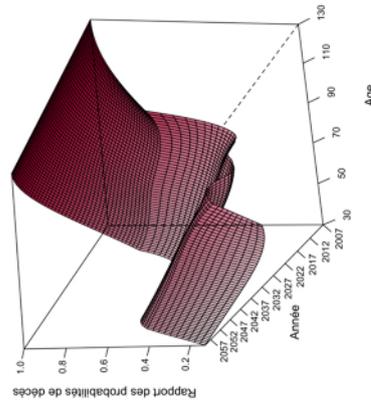
(d) Méthode 4



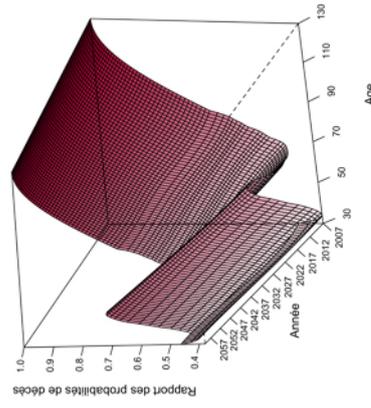
(e) Méthode 1



(f) Méthode 2



(g) Méthode 3



(h) Méthode 4

FIGURE 6: Comparaison de la mortalité masculine et féminine.

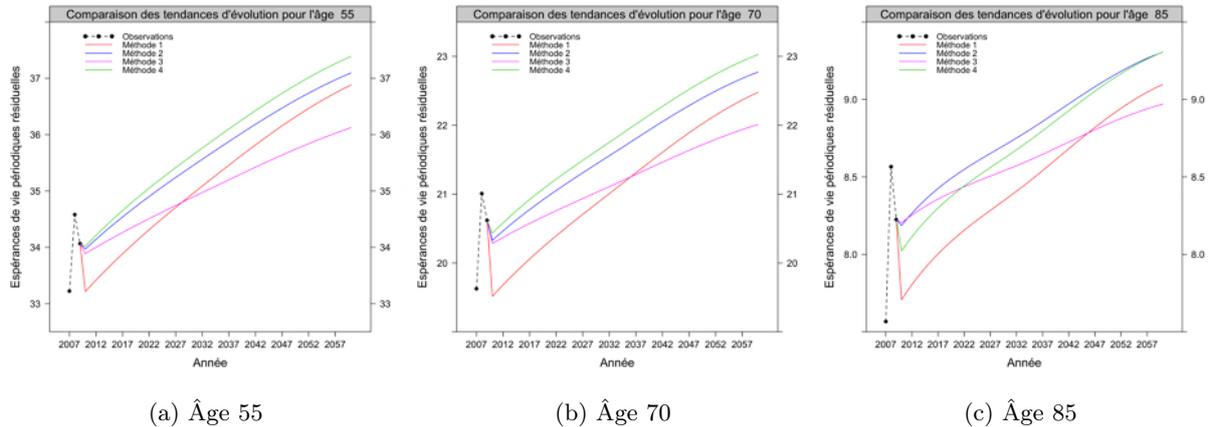


FIGURE 7: Comparaison des tendances d'évolution des espérances de vie périodiques résiduelles par âge, population masculine

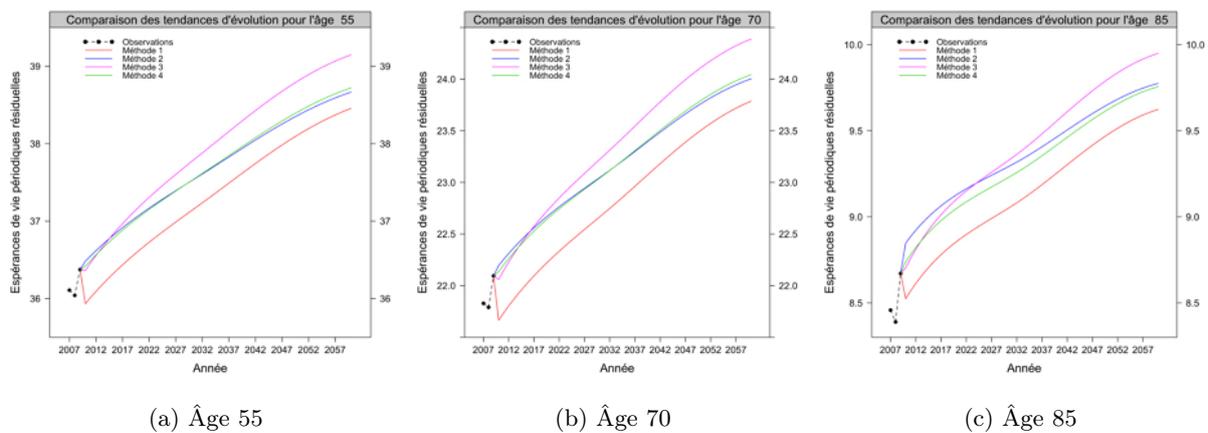


FIGURE 8: Comparaison des tendances d'évolution des espérances de vie résiduelles partielles par âge, population féminine

6.3 Les intervalles de confiances, erreurs relatives et risque systématique d'estimation des paramètres

Pour terminer ces éléments de comparaisons, nous nous intéressons au degré de précision de l'ajustement. Cette notion est appréhendée par les intervalles de confiances sur les décès, espérances de vie et autres quantités relatives à la durée de vie ajustée.

6.3.1 Intervalles de confiances ponctuels sur les décès et erreur relative

Les Figures 9 et 10 comparent les décès observés et ajustés pour la plage d'âges 30-95 et les années 2007 – 2009 pour la population masculine et féminine respectivement. En addition, nous avons ajoutés les intervalles de confiances ponctuels à 95 % pour les taux de décès ajustés.

Nous constatons que les décès observés se situent à l'intérieur des intervalles de confiance théoriques à 95 % sur l'ensemble de la plage d'âge considérée, signe d'une représentation correcte de la réalité par les tables d'expériences masculine et féminine. Néanmoins, quelles que soient les deux populations, la méthode 1 sous-estime les décès pour la plage d'âges 30 – 60 et les sur-estime fortement après 60 ans. A l'opposé, les méthodes 2, 3 et 4 ajustent relativement correctement les décès sur cette plage d'âges pour les deux populations.

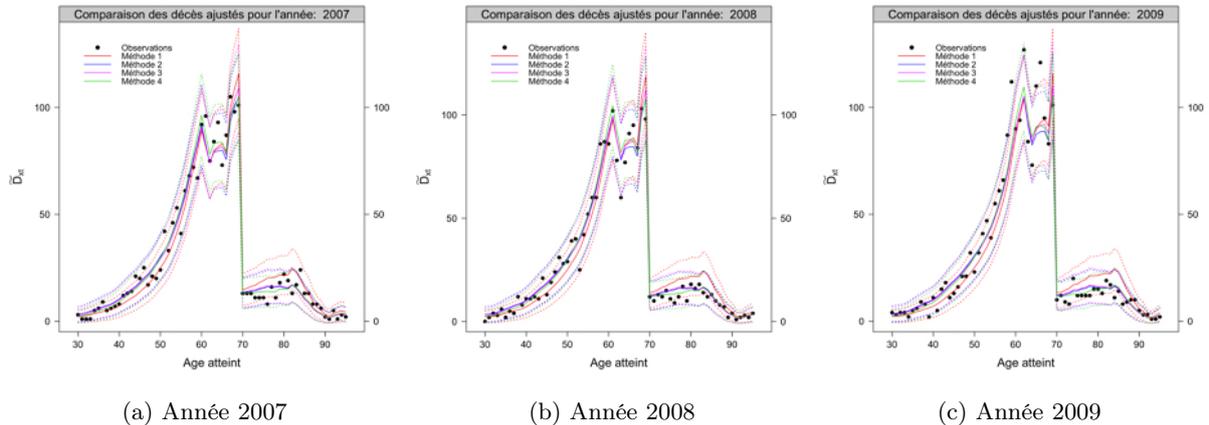


FIGURE 9: Comparaison des décès observés et ajustés, population masculine.

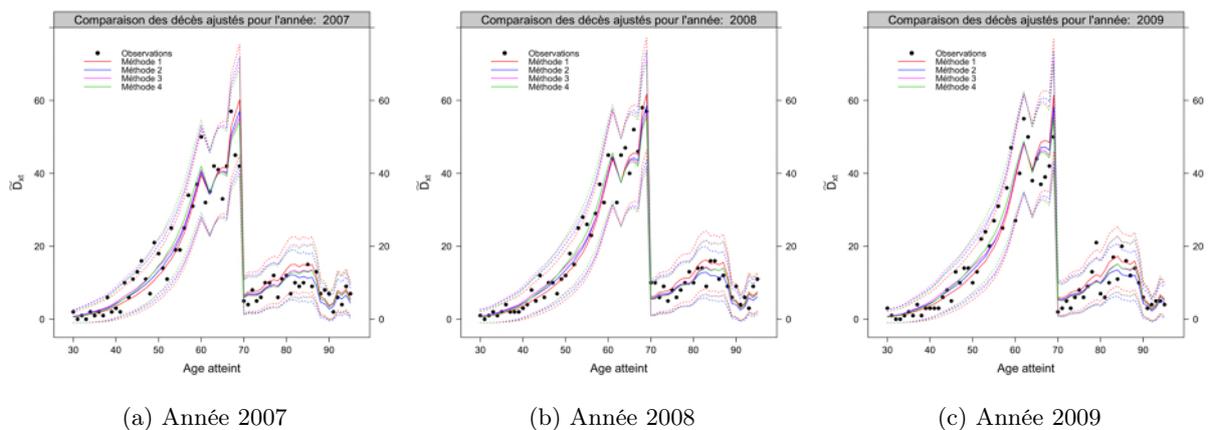


FIGURE 10: Comparaison des décès observés et ajustés, population féminine.

La Table 4 présente l'erreur relative globale sur l'ensemble des âges. Nous observons que l'imprécision est importante, de l'ordre de 45 % et 25 % pour la population masculine et féminine. Par ailleurs, la méthode 4 donne la plus faible erreur relative moyenne.

	Méthode 1	Méthode 2	Méthode 3	Méthode 4
Pop. Masculine	44.91	44.66	44.06	43.02
Pop. Féminine	25.24	26.66	29.26	25.50

TABLE 4: Erreur relative moyenne sur les décès observés (en %).

6.3.2 Intervalles de confiances et dispersion relative sur les espérances de vie ajustées

On souhaite, ici, obtenir les intervalles de confiances ponctuels afin d'encadrer les espérances de vie générationnelles partielles et mesurer la dispersion relative associée à chaque méthode. On ré-échantillonne les taux de décès bruts à partir de la méthode de simulation directe des taux. Pour chaque simulation $k \in [1, 1000]$ et pour chaque approche, on en déduit $q_x^k(t)$ et on estime $\tilde{q}_x^k(t)$.

La Figure 11 présente les probabilités de décès ajustées issues de simulations $\tilde{q}_x^k(t)$ pour chaque méthode sur l'année 2007 et sur plage d'âges $[60 - 80]$ pour la population masculine et féminine.

Ces différents taux ajustés représentent la dispersion des taux de décès ajustés au titre du risque systématique sur l'estimation des paramètres compte tenu des fluctuations d'échantillonnage. On quantifie cette dispersion en calculant une approximation des coefficients de variations des taux de décès, présentées Figures 12a à 12h.

D'une manière générale, l'ampleur de la dispersion augmente avec la complexité des méthodes. Néanmoins, les coefficients de variation varient suivant les méthodes, avec l'année calendaire et l'âge atteint. La méthode 1 à un paramètre entraîne une variation identique selon les années calendaires et les âges jusqu'à 85 ans, puis diminue au fur et à mesure que l'on se rapproche de l'âge 130 de fin de vie induit par le modèle de fermeture, Figures 12a et 12e. Il apparaît avec les méthodes 2, 3 et 4 que la valeur des coefficients de variation a tendance à augmenter avec les années calendaires (pour un âge donné), ce qui est cohérent avec la baisse des taux de décès à un âge donné au fil des années, compte tenu de l'amélioration de l'espérance de vie.

Cette augmentation avec l'année calendaire est beaucoup plus marquée pour la méthode 3, Figures 12c et 12g, que pour les méthodes 2, Figures 12b et 12f et 4, Figures 12d et 12h. De plus, malgré l'augmentation des fluctuations d'échantillonnage lorsque l'âge augmente, on observe une diminution de la valeur des coefficients jusqu'à 70 ans environ. Cette baisse est liée à la hausse des taux de décès au fil des âges. À partir de 70 ans toutefois, les coefficients re-augmentent, ce qui signifie que l'effet lié à la hausse des taux de décès devient plus faible que l'effet lié à la hausse des fluctuations d'échantillonnage dû à la baisse de l'exposition, cf [Kamega and Planchet \(2013\)](#).

Enfin, on encadre les espérances de vie périodiques résiduelles issues des simulations pour les quatre méthodes.

Les Figures 13 et 14 présentes les 2.5 %, 50 % et 97.5 % quantiles des espérances de vie périodiques résiduelles issues des simulations pour la population masculine et féminine respectivement.

Le niveau et l'évolution de la dispersion des espérances de vie périodiques résiduelles simulées sont très différents d'une méthode à une autre que ce soit pour la population masculine ou féminine. La méthode 1, à un paramètre, présente la plus faible dispersion. A l'opposé, le modèle linéaire généralisé de Poisson, méthode 3, entraîne la plus forte variation.

Enfin, on observe que la pente des quantiles à 50 % des espérances de vie périodiques résiduelles simulées (à chaque âge) est supérieure pour la population masculine que pour la population féminine pour les méthodes 1, 2 et 4 indiquant une accélération de l'amélioration de la mortalité plus importante chez les hommes que chez les femmes. A contrario, la méthode 3 entraîne une amélioration plus importante pour la population féminine.

La dispersion relative des espérances de vie périodiques résiduelles issues des simulations est présentée en Figures 15 et 16 pour la population masculine et féminine respectivement.

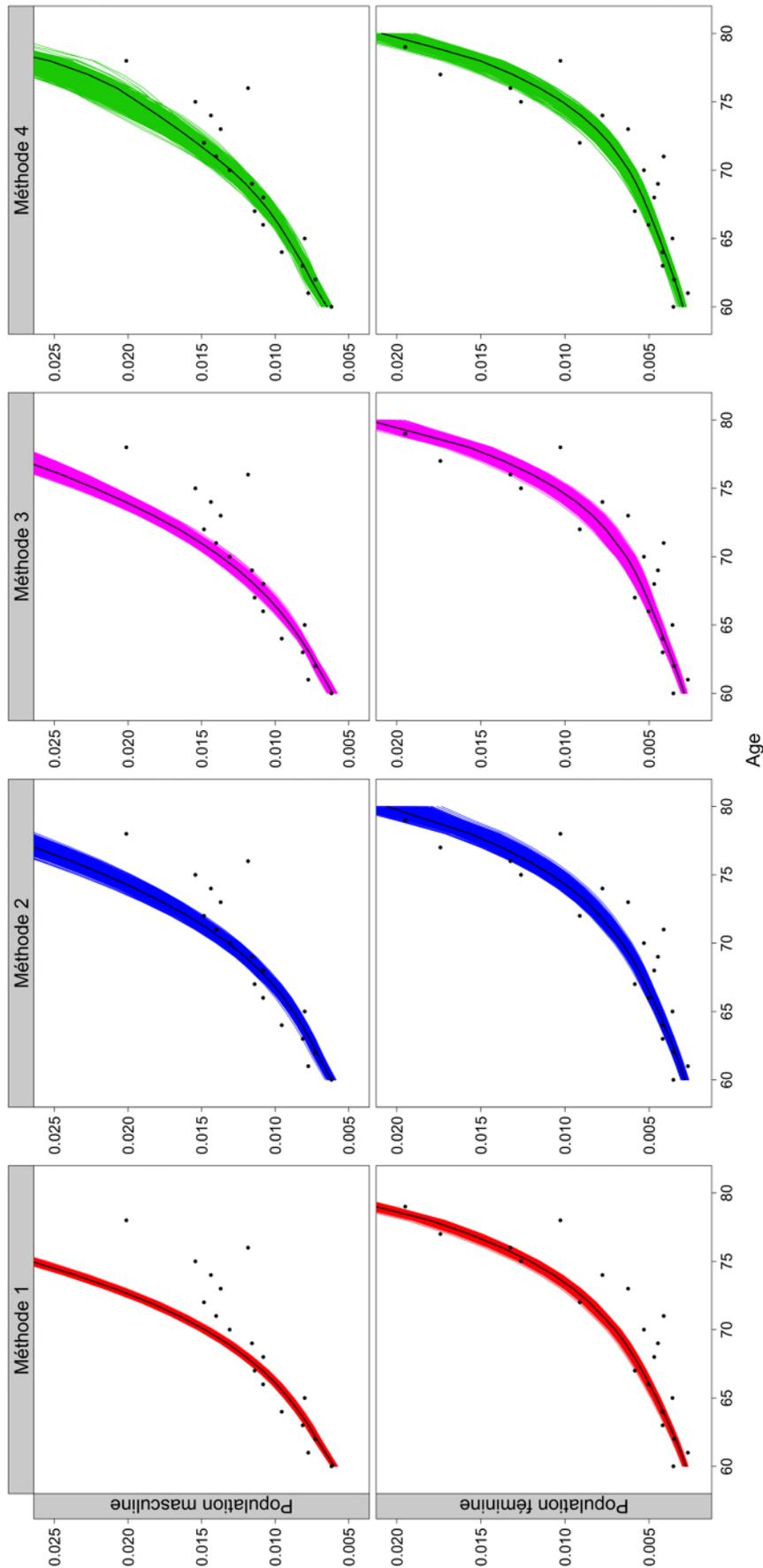
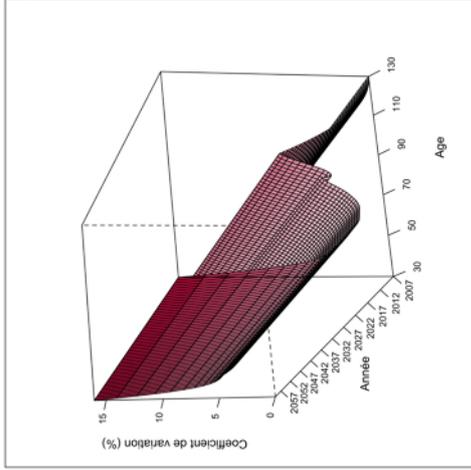
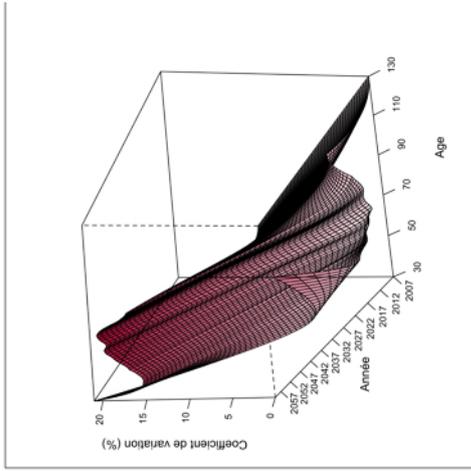


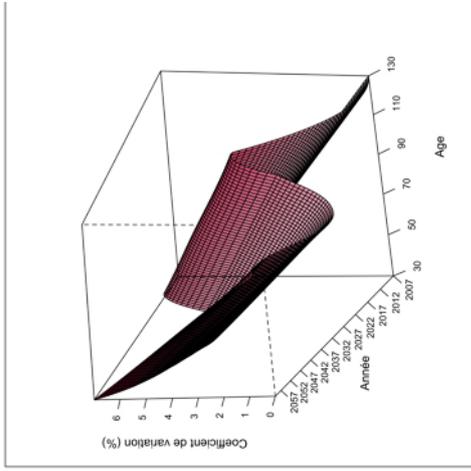
FIGURE 11: Probabilités de décès ajustées issues des 1000 simulations sur l'année 2007 et sur plage d'âges [60 – 80] pour la population masculine et féminine.



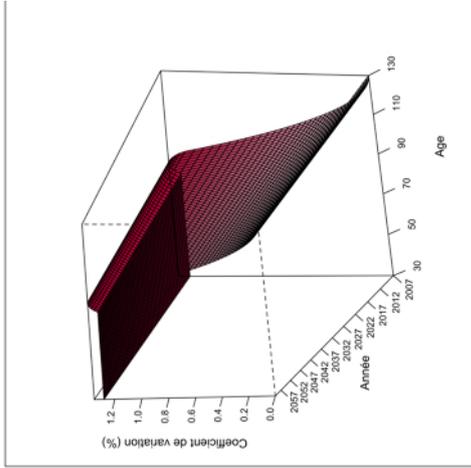
(a) Méthode 1, pop. masculine.



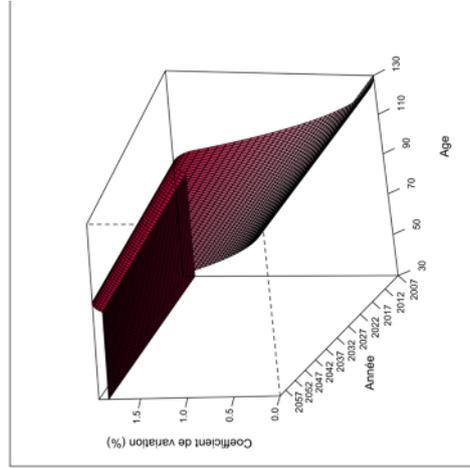
(b) Méthode 2, pop. masculine.



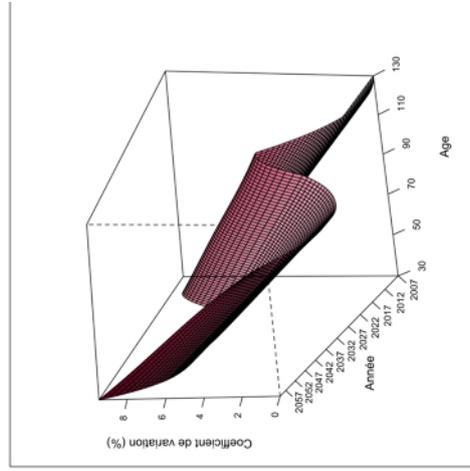
(c) Méthode 3, pop. masculine.



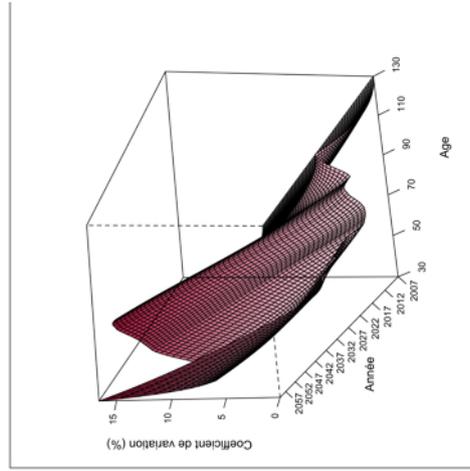
(d) Méthode 4, pop. masculine.



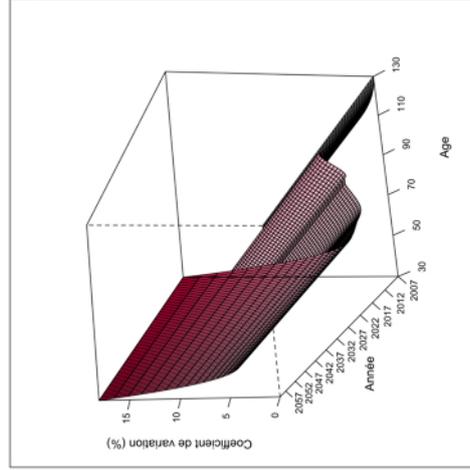
(e) Méthode 1, pop. féminine.



(f) Méthode 2, pop. féminine.



(g) Méthode 3, pop. féminine.



(h) Méthode 4, pop. féminine.

FIGURE 12: Coefficient de variation (en %) sur la plage d'âges [30 – 130] et sur les années 2007 – 2060 pour la population masculine et féminine.

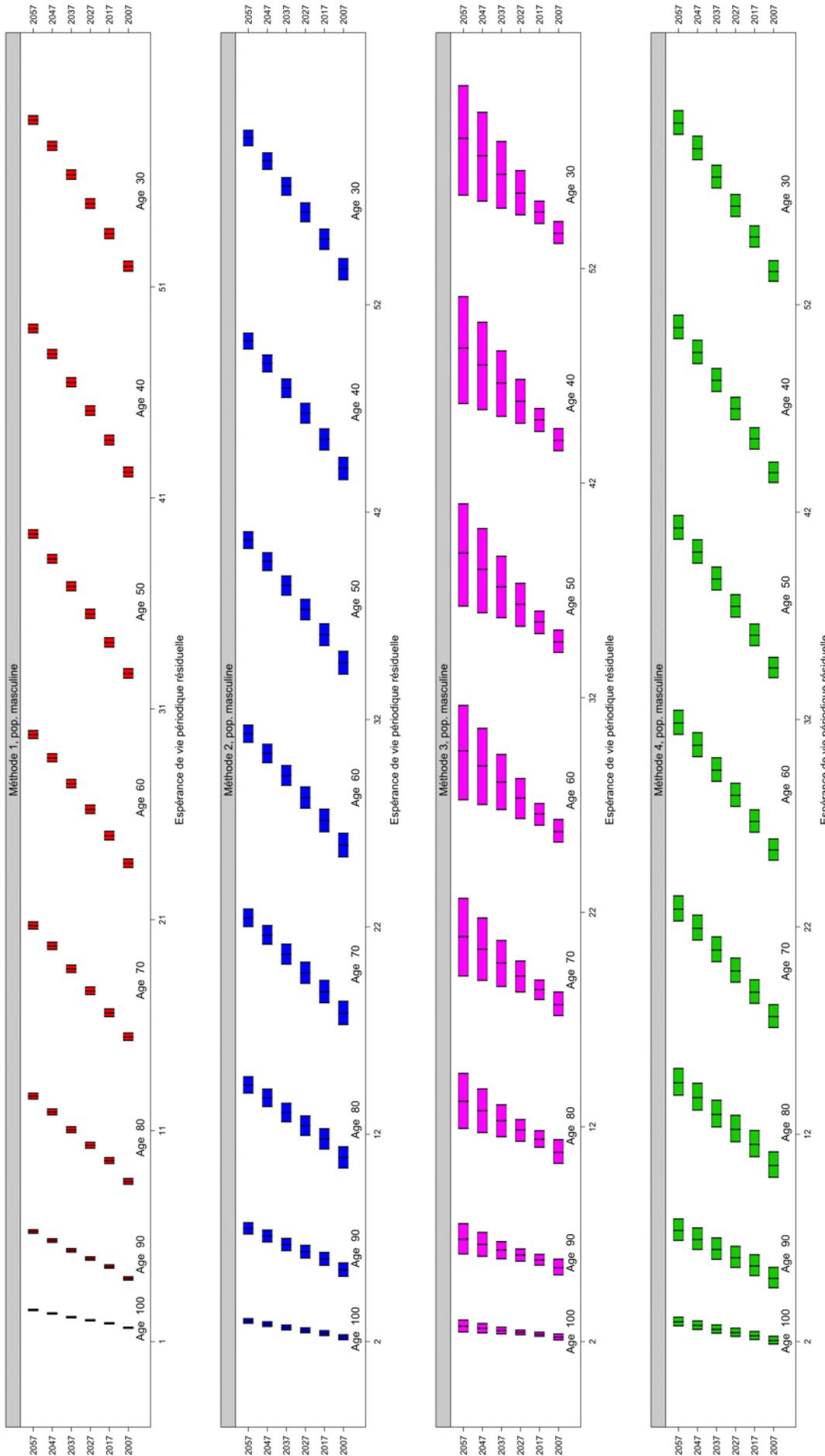


FIGURE 13: 2.5 %, 50 % et 97.5 % quantiles des espérances de vie périodiques résiduelles issues des simulations pour la population masculine.

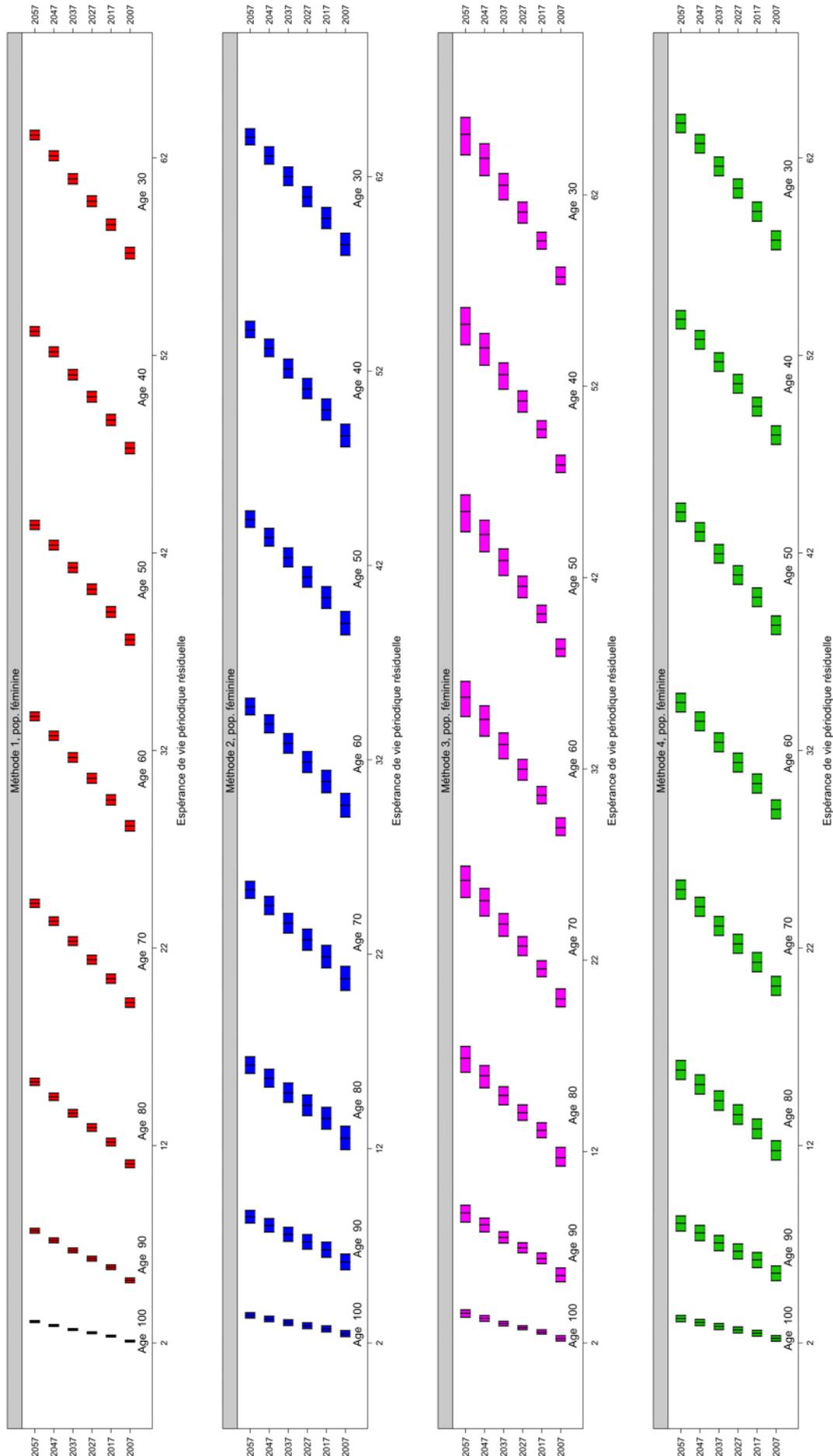


FIGURE 14: 2.5 %, 50 % et 97.5 % quantiles des espérances de vie périodiques résiduelles issues des simulations pour la population féminine.

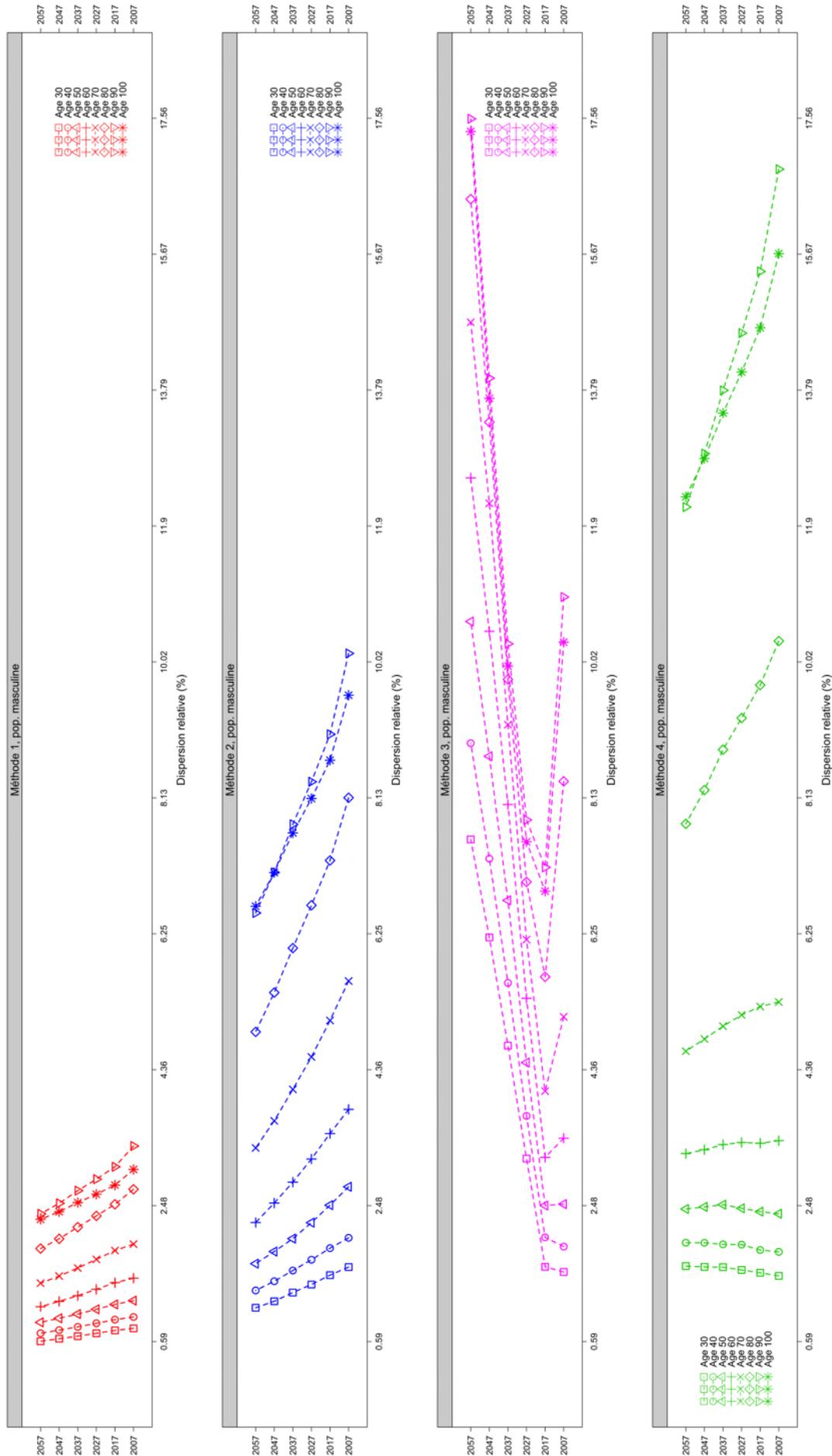


FIGURE 15: Dispersion relative (en %) des espérances de vie périodiques résiduelles issues des simulations pour la population masculine.

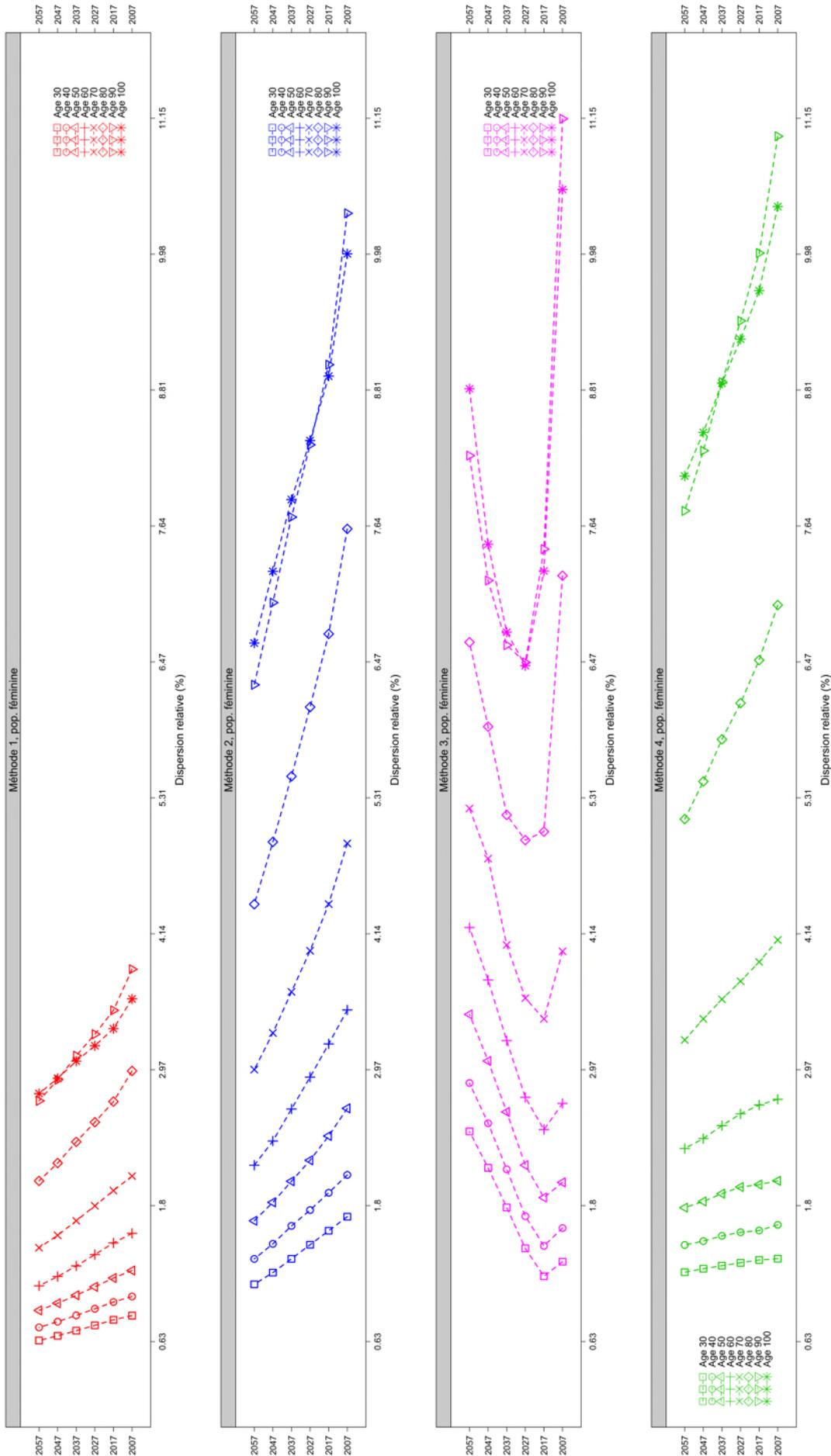


FIGURE 16: Dispersion relative (en %) des espérances de vie périodiques résiduelles issues des simulations pour la population féminine.

Conformément aux remarques faites précédemment, le degré et le profil de la dispersion relative varient énormément d'une méthode à une autre que ce soit pour la population masculine ou féminine. La méthode 1 présente le degré de dispersion relative le plus faible, alors que la méthode 3 entraîne la plus large dispersion. Les méthodes 1, 2 et 4 induisent une diminution de la dispersion avec l'année calendaire alors que la méthode 3 conduit à une forte augmentation au fil du temps.

7 Conclusion

Cette note de travail présente les critères quantitatifs et qualitatifs envisagés pour évaluer la pertinence des approches concernant la construction de tables de mortalité de place.

Les divergences entre les approches sont appréhendées sur trois niveaux. celles-ci concernent l'ajustement de la mortalité passée, la tendance d'évolution de la mortalité et autres quantités relatives à la durée de vie ajustée ainsi que les intervalles de confiances sur les décès, et la dispersion des espérances de vie.

Références

- Cairns, A. J. G., Blake, D., and Dowd, K. (2006). Pricing death : Frameworks for the valuation and securization of mortality risk. *ASTIN Bulletin*, **36**, 79–120.
- Debón, A., Montes, F., and Sala, R. (2006). A comparison of nonparametric methods in the graduation of mortality : Application to data from the Valencia region (Spain). *International statistical Review*, **74**(2), 215–233.
- Felipe, A., Guillén, M., and Pérez-Marín, A. (2002). Recent mortality trends in the Spanish population. *British Actuarial Journal*, **8**(4), 757–786.
- Forfar, D., McCutcheon, J., and Wilkie, A. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, **115**(1), 1–459.
- Kamega, A. and Planchet, F. (2013). Construction de tables de mortalité prospectives sur un groupe restreint : Mesure du risque d'estimation. *Bulletin Français d'Actuariat*, **13**(25).
- Liddell, F. D. K. (1984). Simple exact analysis of the standardised mortality ratio. *Journal of Epidemiology and Community Health*, **38**, 85–88.
- Pitacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press.
- Planchet, F. and Thérond, P. (2011). *Modélisation statistique des phénomènes de durée - Applications actuarielles*. Assurance Audit Actuariat. Economica Paris.